



# Optimal subsets in the stability regions of multistep methods

Lajos Lóczi<sup>1,2</sup> 

Received: 24 January 2019 / Accepted: 3 July 2019 / Published online: 22 July 2019  
© The Author(s) 2019

## Abstract

In this work, we study the stability regions of linear multistep or multiderivative multistep methods for initial value problems by using techniques that are straightforward to implement in modern computer algebra systems. In many applications, one is interested in (i) checking whether a given subset of the complex plane (e.g., a sector, disk, or parabola) is included in the stability region of the numerical method, (ii) finding the largest subset of a certain shape contained in the stability region of a given method, or (iii) finding the numerical method in a parametric family of multistep methods whose stability region contains the largest subset of a given shape. First, we describe a simple procedure to exactly calculate the stability angle  $\alpha$  in the definition of  $A(\alpha)$ -stability by representing the root locus curve of the multistep method as an implicit algebraic curve. As an illustration, we consider two finite families of implicit multistep methods. We exactly compute the stability angles for the  $k$ -step BDF methods ( $3 \leq k \leq 6$ ) and discover that the values of  $\tan(\alpha)$  are surprisingly simple algebraic numbers of degree 2, 2, 4, and 2, respectively. In contrast, the corresponding values of  $\tan(\alpha)$  for the  $k$ -step second-derivative multistep methods of Enright ( $3 \leq k \leq 7$ ) are much more complicated; the smallest algebraic degree here is 22. Next, we determine the exact value of the stability radius in the BDF family for each  $3 \leq k \leq 6$ , that is, the radius of the largest disk in the left half of the complex plane, symmetric with respect to the real axis, touching the imaginary axis

---

The project is supported by the Hungarian Government and co-financed by the European Social Fund, EFOP-3.6.3-VEKOP-16-2017-00001: Talent Management in Autonomous Vehicle Control Technologies.

---

✉ Lajos Lóczi  
[LLoczi@inf.elte.hu](mailto:LLoczi@inf.elte.hu)

<sup>1</sup> Department of Numerical Analysis, Faculty of Informatics, ELTE Eötvös Loránd University, Budapest, Hungary

<sup>2</sup> Department of Differential Equations, Budapest University of Technology and Economics, Budapest, Hungary

and lying in the stability region of the corresponding method. These radii turn out to be algebraic numbers of degree 2, 3, 5, and 5, respectively. Finally, we demonstrate how some Schur–Cohn-type theorems of recursive nature and not relying on the root locus curve method can be used to exactly solve some optimization problems within infinite parametric families of multistep methods. As an example, we choose a two-parameter family of implicit-explicit (IMEX) methods: we identify the unique method having the largest stability angle in the family, then we find the unique method in the same family whose stability region contains the largest parabola.

**Keywords** Multistep methods · Optimal subsets in the stability region · Exact optimization · Stability angle · Stability radius

## 1 Introduction

In the stability theory of one-step or multistep methods for initial value problems, one is often interested in various geometric properties of the stability region  $\mathcal{S} \subset \mathbb{C}$  of the method. In this work, we study the shape of the stability region of linear multistep methods (LMMs) or multiderivative multistep methods (also known as generalized LMMs) as follows.

Suppose we are given

- a) a stability region  $\mathcal{S}$  or
- b) a family of stability regions  $\mathcal{S}_\beta$  parametrized by some  $\beta \in \mathbb{R}^d$

and a family of subsets of  $\mathbb{C}$ , denoted by  $\mathfrak{F}$ . Due to their relevance in applications, we will consider the following three classes:

- $\mathfrak{F} = \mathfrak{F}_\alpha^{\text{sect}}$  is the family of infinite sectors in the left half of  $\mathbb{C}$ , with vertex at the origin, symmetric about the negative real axis, and parametrized by the sector angle  $\alpha \in (0, \pi/2)$ .
- $\mathfrak{F} = \mathfrak{F}_r^{\text{disk}}$  is the family of disks in the left half of  $\mathbb{C}$ , symmetric with respect to the real axis, touching the imaginary axis, and parametrized by the disk radius  $r > 0$ .
- $\mathfrak{F} = \mathfrak{F}_m^{\text{para}}$  is the family of parabolas in the left half of  $\mathbb{C}$ , symmetric with respect to the real axis, touching the imaginary axis, and parametrized by some  $m > 0$ .

Our goal is to find the set  $H \in \mathfrak{F}$  with the largest parameter ( $\alpha$ ,  $r$ , or  $m$ ) such that

- $H \subset \mathcal{S}$  in case a;
- $H \subset \mathcal{S}_{\beta_{\text{opt}}}$  for some stability region in the family in case b, but  $H \not\subset \mathcal{S}_\beta$  for  $\beta \neq \beta_{\text{opt}}$ .

We will present some tools to handle these shape optimization questions and, as an illustration, exactly solve some of them by using *Mathematica* version 11 in the BDF (backward differentiation formula) and Enright families (as LMMs and multiderivative multistep methods, respectively), and in an infinite family of IMEX methods with  $d = 2$  parameters.

## 1.1 Motivation and main results

**A** When solving stiff ordinary differential equations, one desirable property of the numerical method is  $A$ -stability: a method is  $A$ -stable if the closed left half-plane  $\{z \in \mathbb{C} : \operatorname{Re}(z) \leq 0\}$  belongs to  $\mathcal{S}$ . Many useful methods are not  $A$ -stable, still,  $\mathcal{S}$  contains a sufficiently large infinite sector in the left half-plane with vertex at the origin and symmetric about the negative real axis. This leads to the notion of  $A(\alpha)$ -stability: a method is  $A(\alpha)$ -stable with some  $0 < \alpha < \pi/2$  if

$$\{z \in \mathbb{C} \setminus \{0\} : |\arg(-z)| \leq \alpha\} \subset \mathcal{S}, \quad (1)$$

where the argument of a non-zero complex number satisfies  $-\pi < \arg \leq \pi$ . The largest  $0 < \alpha < \pi/2$  such that (1) holds is referred to as the stability angle of the method [17]. Various other stability concepts—such as  $A(0)$ -stability,  $A_0$ -stability,  $A$ -stability, stiff stability, or asymptotic  $A(\alpha)$ -stability—have also been defined, and theorems are devised to test whether a given multistep method is stable in one of the above senses (see, for example, [3, 9, 11, 20–23, 26, 34, 40]). There are various techniques to test  $A(\alpha)$ -stability for a given  $\alpha$  value. In [3], for example, the sector on the left-hand side of (1) is decomposed into an infinite union of disks, and a bijection between each disk and the left half-plane is established via fractional linear transformations to employ a Routh–Hurwitz-type criterion. Another way of studying  $A(\alpha)$ -stability is to consider the root locus curve (RLC) of the multistep method [17]. Based on the RLC and some theorems from complex analysis, [38] presents a criterion for a LMM to be  $A(\alpha)$ -stable for a given  $\alpha$ ; the stability angle is then obtained as the solution of an optimization problem involving Chebyshev polynomials. The procedure in [38] is formulated only for LMMs but not for multiderivative multistep methods.

The first goal of the present work is to describe an elementary approach to exactly determine the stability angle of a LMM or multiderivative multistep method: by eliminating the complex exponential function from the RLC and using a tangency condition, a system of polynomial equations in two variables is set up whose solution yields the stability angle. This process is easily implemented in computer algebra systems. As an illustration, we consider two finite families: the BDF methods [13, 17, 38] as LMMs and the second-derivative multistep methods of Enright [6, 10, 17]. With  $\alpha_k^{\text{BDF}}$  denoting the stability angle of the  $k$ -step BDF method for  $3 \leq k \leq 6$ , we show that  $\tan(\alpha_k^{\text{BDF}})$  is an unexpectedly simple algebraic number, having degree 2 for  $k \in \{3, 4, 6\}$  and degree 4 for  $k = 5$  (see Table 1). For the  $k$ -step Enright methods with  $3 \leq k \leq 7$ , the corresponding constants  $\tan(\alpha_k^{\text{Enr}})$  (with approximate values listed in Table 2) are much more complicated algebraic numbers of increasing degree (starting with 22). As far as we know, exact values  $\alpha \in (0, \pi/2)$  for the stability angles of multistep methods were not presented earlier in the literature.

**Remark 1.1** The  $k$ -step BDF methods for  $k \in \{1, 2\}$  are  $A$ -stable. For  $k \geq 7$  they are not zero-stable [7, 8, 16], therefore not interesting from a practical point of view.

**Table 1** The exact stability angles  $\alpha_k^{\text{BDF}} = \frac{180}{\pi} \arctan(c_k^{\text{BDF}})$  of the BDF methods expressed in degrees

$k$	$c_k^{\text{BDF}}$	Approximate value of $\alpha_k^{\text{BDF}}$
3	$\frac{329\sqrt{\frac{7}{5}}}{27}$	86.032366860211647332°
4	$\frac{699\sqrt{\frac{3}{2}}}{256}$	73.351670474578482110°
5	$\frac{1326107429}{25} \sqrt{\frac{62}{53860574450525125+1194498034900685\sqrt{2033}}}$	51.839755836049910391°
6	$\frac{45503}{10125\sqrt{195}}$	17.839777792245700101°

**Remark 1.2** In [38, Table 1], one finds some approximate values for the BDF stability angles; however, some of these values are not correct. The  $k = 3$  value is wrong because the polynomial  $R_3$  is not computed properly (a factor 2 is missing, the correct form would have been  $R_3(x) = -12(x - 1)^2(4x - 1)$ ). The approximate values for  $k = 4$  and  $k = 5$  given in [38, Table 1] are correct (up to the given precision). The value for  $k = 6$  is again incorrect because an error was committed in the minimization process. If the optimization in [38, Section 3] is carried out exactly with the correct  $R_j$  polynomials, we recover the stability angle values in our Table 1. The errors in [38, Table 1] propagated in the literature (see, for example, [33, p. 242]). As a consequence, some works that appeared in the current millennium also contain the erroneous angles. In [17, Chapter V.2, (2.7)], the correct approximate values are presented.

**Remark 1.3** At the time of writing this document, we learned (through personal communication) that [1] also contains the exact stability angles for the BDF methods with  $3 \leq k \leq 6$  steps: although they use a different technique to derive the results and the arcsin function to express the final constants, the values given in [1] and our Table 1 are the same. Notice, however, that the stability angle for  $k = 5$  given in [1] has a slightly more complicated structure than the value in our Table 1.

**Remark 1.4** The  $k$ -step Enright methods are  $A$ -stable again for  $k \in \{1, 2\}$  (see [17]) and unstable for  $k \geq 8$ . More precisely, [11] proves that these methods are not  $A_0$ -stable for  $k \geq 8$ ; hence, they cannot be stiffly stable either (see [23, Theorem 3])

**Table 2** Stability angles  $\alpha_k^{\text{Enr}} = \frac{180}{\pi} \arctan(c_k^{\text{Enr}})$  of the Enright methods expressed in degrees

$k$	Approximate value of $c_k^{\text{Enr}}$	Approximate value of $\alpha_k^{\text{Enr}}$
3	27.056933440109472532101963	87.8833627693413031369003498°
4	7.1406622283653916403051061	82.0279713768712835947479188°
5	3.2907685080317853840110455	73.0970020659749082763655203°
6	1.7285146253131256601603521	59.9492702555400766770433070°
7	0.7703217281441388675578954	37.6078417405752150238159031°

(cf. [22, 26]). However, in [17, Chapter V.3, p. 276, Exercise 2], the stiff instability of the Enright formulae for  $k \geq 8$  is still mentioned as an open problem.

**B** The stability radius of a multistep method is the largest number  $r > 0$  such that the inclusion

$$\{z \in \mathbb{C} : |z + r| \leq r\} \subset \mathcal{S}$$

holds. The stability radius plays an important role when analyzing the boundedness properties of multistep methods. For example, it has been proved [42, Theorem 3.1] that this radius is the largest step-size coefficient for linear boundedness of a LMM satisfying some natural assumptions.

*Remark 1.5* For LMMs (and for more general methods as well), various other step-size coefficients have been introduced in the context of linear or non-linear problems. These coefficients govern the largest allowable step-size guaranteeing certain monotonicity or boundedness properties of the LMM, including the TVD and SSP properties [15]. These properties are relevant, for example, in the time integration of method-of-lines semi-discretizations of hyperbolic conservation laws [19, 35, 41].

*Remark 1.6* In [31], the largest inscribed and smallest circumscribed (semi)disks are computed for certain one-step methods.

The second goal of the present work is to compute the stability radius for some multistep methods. We will achieve this by using again the algebraic form of the RLCs. Table 3 contains the exact values in the BDF family for  $3 \leq k \leq 6$ .

**C** The RLC, as the graph of a  $[0, 2\pi] \rightarrow \mathbb{C}$  function (or a union of such functions for generalized LMMs), yields information about the boundary of the stability region,  $\partial\mathcal{S}$ . It is known, however, that in general the RLC does not coincide with  $\partial\mathcal{S}$  (see Fig. 3). This does not pose a problem when a fixed multistep method is considered—one can evaluate the roots of the characteristic polynomial at finitely many test points sampled from different components of  $\mathbb{C}$  determined by the RLC to see which component belongs to  $\mathcal{S}$  and which one to  $\mathbb{C} \setminus \mathcal{S}$ . But when working with parametric families of multistep methods, the precise identification of the stability region boundaries or components can become challenging with the RLC

**Table 3** The exact stability radii  $r_k^{\text{BDF}}$  of the BDF methods

$k$	$r_3^{\text{BDF}}$ is equal to / $r_{4,5,6}^{\text{BDF}}$ is a root of the polynomial	Approximate value of $r_k^{\text{BDF}}$
3	$(17 + 8\sqrt{10})/6$	7.049703546891172
4	$\{18432, 2172, -100855, -114975\}$	2.727199466336645
5	$\{2944512000, 260854387200, 679386763440, 266052478296, -1280160594125, -1354065829875\}$	1.357947301777465
6	$\{141717600000, 558150393600, 1112790780640, 948530730784, -119637602525, -488414721375\}$	0.559931687924882

method. One can overcome this difficulty for example by invoking a reduction process, the Schur–Cohn reduction, formulated in, e.g., [37]. Instead of using auxiliary fractional linear transformations and applying Routh–Hurwitz-type criteria [28, 36] as mentioned above, these Schur–Cohn-type theorems in [37] are directly tailored to the context of multistep methods to locate the roots of the characteristic polynomials with respect to the unit disk.

The third goal of the present work is to demonstrate the effectiveness of the Schur–Cohn reduction when we solve two optimization case studies in a family of implicit-explicit (IMEX) multistep methods taken from [18]. On the one hand, we find the method in the IMEX family that has the largest stability angle, that is, the method whose stability region contains the largest sector (see our Theorem 5.3). On the other hand, we illustrate the versatility of the reduction technique by also finding the method whose stability region contains the largest parabola (see Theorem 6.1); the inclusion of a parabola-shaped region in  $\mathcal{S}$  is relevant when studying semi-discretizations of certain partial differential equations (PDEs) of advection-reaction-diffusion type [5, 18, 28]. The chosen IMEX family is described by two real parameters, and the corresponding characteristic polynomial is cubic. The Schur–Cohn reduction process recursively decreases the degree of the characteristic polynomial, so instead of analyzing the roots of high-degree polynomials, we finally need to check polynomial inequalities in the parameters present in the coefficients of the original polynomial. Besides the two real parameters, two complex variables are involved in our calculations—the non-trivial interplay between these six real variables determines the optimum in both cases. We emphasize that we solve the optimization problem exactly, and RLCs are not relied on in the rigorous part of the proofs (only when setting up conjectures about the optimal values).

*Remark 1.7* The Schur–Cohn reduction is also used in [25] to explore certain properties of a discrete parametric family of multistep methods. Conditions for disk or segment inclusions in the stability regions of a two-parameter family of multistep methods are formulated in [39]. Optimality questions about the size and shape of the stability regions of one-step or multistep methods are investigated in detail in [27]. Properties of optimal stability polynomials and stability region optimization in parametric families of one-step methods are discussed, for example, in [29, 30].

## 1.2 Structure of the paper

In Section 2.1, we introduce some notation. In Sections 2.2–2.3, we review the Schur–Cohn reduction and the definition of the stability region of a multistep method. In Sections 2.4–2.5, the definition of the root locus curve is recalled in two special cases: for linear multistep methods and for second-derivative multistep methods. Here, we consider the BDF and Enright families as concrete examples.

Regarding the new results, a simple algebraic technique is described in Section 3.1 to exactly compute the stability angle of a linear multistep or multiderivative multistep method. Stability angles for the BDF and Enright families are tabulated in Sections 3.2–3.3. In Section 4, we exactly compute the stability radii in the BDF family by using the same approach. In Section 5, we first describe a two-parameter family

of IMEX multistep methods, in which we determine the unique method with the largest stability angle, then, in Section 6, the unique method whose stability region contains the largest parabola. The techniques in Sections 5–6 do not rely on root locus curves but use the Schur–Cohn reduction instead; the full proofs are deferred to Appendices A and B.

## 2 Preliminaries

### 2.1 Notation

The set of natural numbers  $\{0, 1, \dots\}$  is denoted by  $\mathbb{N}$ . For  $z \in \mathbb{C}$ ,  $\operatorname{Re}(z)$ ,  $\operatorname{Im}(z)$ , and  $\bar{z}$  denote the real and imaginary parts and the conjugate of  $z$ , respectively, and  $i$  is the imaginary unit. The boundary of a (possibly unbounded) set  $H \subset \mathbb{C}$  is  $\partial H \subset \mathbb{C}$ . When describing certain algebraic numbers of higher degree, a polynomial  $\sum_{j=0}^n a_j x^j$  with  $a_j \in \mathbb{Z}$ ,  $a_n \neq 0$  and  $n \geq 3$  will be represented simply by its coefficient list  $\{a_n, a_{n-1}, \dots, a_0\}$ . For a polynomial  $Q(z) = \sum_{j=0}^n a_j z^j$  with  $0 \leq n \in \mathbb{N}$ ,  $a_j \in \mathbb{C}$  ( $0 \leq j \leq n$ ), and  $a_n \neq 0$ , we denote its degree, leading coefficient and constant coefficient by  $\deg Q = n$ ,  $\operatorname{lc} Q = a_n$ , and  $\operatorname{cc} Q = a_0$ . The acronyms RLC and LMM stand for root locus curve and linear multistep method, respectively.

### 2.2 The Schur–Cohn reduction

In the rest of this section, we assume that  $Q$  is a univariate polynomial with  $\deg Q \geq 1$ , and follow the terminology of [37]—we have explicitly added the  $\deg Q \geq 1$  condition, being implicit in [37]. We say that:

- $Q$  is a *Schur polynomial*,  $Q \in \mathbf{Sch}$ , if its roots lie in the open unit disk.
- $Q$  is a *von Neumann polynomial*,  $Q \in \mathbf{vN}$ , if its roots lie in the closed unit disk.
- $Q$  is a *simple von Neumann polynomial*,  $Q \in \mathbf{svN}$ , if  $Q \in \mathbf{vN}$  and roots with modulus 1 are simple.

*Remark 2.1* The class **Sch** is referred to as strongly stable polynomials in [4, p. 345].

*Remark 2.2* The property  $Q \in \mathbf{svN}$  is often expressed by saying that  $Q$  satisfies the root condition.

The reduced polynomial of  $Q(z) = \sum_{j=0}^n a_j z^j$  is defined as

$$Q^r(z) := \frac{\overline{a_n} \cdot \left( \sum_{j=0}^n a_j z^j \right) - a_0 \cdot \left( \sum_{j=0}^n \overline{a_{n-j}} z^j \right)}{z} = \sum_{j=1}^n (\overline{a_n} \cdot a_j - a_0 \cdot \overline{a_{n-j}}) z^{j-1},$$

so we have  $\deg Q^r \leq (\deg Q) - 1$ . When this reduction process is iterated, we write  $Q^{rr}$  for  $(Q^r)^r$ , for example. The following theorems from [37] use the notion of the

reduced polynomial and the derivative to formulate necessary and sufficient conditions for a polynomial to be in the above classes. In all three theorems below, it is assumed that  $\text{lc } Q \neq 0 \neq \text{cc } Q$  and  $\deg Q \geq 2$ .

**Theorem 2.3**  $Q \in \text{Sch} \Leftrightarrow (|\text{lc } Q| > |\text{cc } Q| \text{ and } Q^r \in \text{Sch}).$

**Theorem 2.4**  $Q \in \mathbf{vN} \Leftrightarrow \text{either } (|\text{lc } Q| > |\text{cc } Q| \text{ and } Q^r \in \mathbf{vN}) \text{ or } (Q^r \equiv 0 \text{ and } Q' \in \mathbf{vN}).$

**Theorem 2.5**  $Q \in \mathbf{svN} \Leftrightarrow \text{either } (|\text{lc } Q| > |\text{cc } Q| \text{ and } Q^r \in \mathbf{svN}) \text{ or } (Q^r \equiv 0 \text{ and } Q' \in \text{Sch}).$

*Remark 2.6* Let us consider the following example when applying the theorems above, e.g., Theorem 2.4. For any  $\lambda > 0$ , we set  $Q_\lambda(z) := z^2 + \lambda iz + 1$ . Then the roots of  $Q_\lambda$  satisfy  $|z_1(\lambda)| < 1 < |z_2(\lambda)|$ , so  $Q_\lambda \notin \mathbf{vN}$ , and  $Q_\lambda^r = 2\lambda i$ . This shows that it can happen that the degree of the original polynomial is  $> 1$ , but its reduced polynomial is a non-zero constant, so the relation  $Q^r \in \mathbf{vN}$  is undefined. In these cases, when  $Q^r$  is a non-zero constant, notice that neither  $|Q^r| < 1$  nor  $|Q^r| = 1$  nor  $|Q^r| > 1$  can help us in general to determine whether  $Q \in \mathbf{vN}$  or not (of course, the other condition  $|\text{lc } Q| > |\text{cc } Q|$  is violated now) (cf. the sentence above [37, Theorem 5.1]).

### 2.3 The stability region of a multistep method

Stability properties of a broad class of numerical methods (including Runge–Kutta methods, linear multistep methods, or multiderivative multistep methods) for solving initial value problems of the form

$$y'(t) = f(t, y(t)), \quad y(t_0) = y_0 \quad (2)$$

can be analyzed by studying the stability region of the method. When an  $s$ -stage  $k$ -step method ( $s \geq 1$ ,  $k \geq 1$  fixed positive integers; for  $k = 1$  we have a one-step method, while for  $k \geq 2$  a multistep method) with constant step size  $h > 0$  is applied to the linear test equation  $y' = \lambda y$  ( $\lambda \in \mathbb{C}$  fixed,  $y(t_0) = y_0$  given), the method yields a numerical solution  $(y_n)_{n \in \mathbb{N}}$  that approximates the exact solution  $y$  at time  $t_n := t_0 + nh$  and satisfies a recurrence relation of the form [27]

$$\begin{cases} \sum_{j=0}^s \sum_{\ell=0}^k a_{j,\ell} \mu^j y_{n+\ell} = 0, & n \in \mathbb{N}, \\ a_{j,\ell} \in \mathbb{R}, \quad \sum_{j=0}^s |a_{j,k}| > 0, \quad \mu := h\lambda. \end{cases} \quad (3)$$

The characteristic polynomial associated with the method takes the form

$$\Phi(\zeta, \mu) := \sum_{j=0}^s \sum_{\ell=0}^k a_{j,\ell} \mu^j \zeta^\ell \quad (\zeta \in \mathbb{C}). \quad (4)$$



With  $\Phi(\cdot, \mu)$  abbreviating the polynomial  $\zeta \mapsto \Phi(\zeta, \mu)$ , the stability region of the method is defined as

$$\mathcal{S} := \{\mu \in \mathbb{C} : \text{the degree of } \Phi(\cdot, \mu) \text{ is exactly } k, \text{ and } \Phi(\cdot, \mu) \in \mathbf{svN}\}. \quad (5)$$

*Remark 2.7* Some other variations of the above definition of the stability region of a multistep method have also been proposed in the literature (see, e.g., [24]). In [4, p. 344], the “open stability region” is defined as the set

$$\{\mu \in \mathbb{C} : \Phi(\cdot, \mu) \in \mathbf{Sch}\},$$

(see also [44, p. 348], [12, p. 452], or [33]). In, e.g., [17, 32], the stability region of the method (3) is defined as

$$\begin{aligned} \{\mu \in \mathbb{C} : \text{all roots } \zeta_j(\mu) \text{ of } \zeta \mapsto \Phi(\zeta, \mu) \text{ satisfy } |\zeta_j(\mu)| \leq 1, \\ \text{and multiple roots satisfy } |\zeta_j(\mu)| < 1\}, \end{aligned} \quad (6)$$

that is, essentially,  $\Phi(\cdot, \mu) \in \mathbf{svN}$ . In [27, Formula (2.5)], the stability region is given by

$$\begin{aligned} \{\mu \in \overline{\mathbb{C}} : \text{roots } \zeta_j \text{ of } \Phi(\zeta, \mu) = 0 \text{ satisfy } |\zeta_j(\mu)| \leq 1, \\ \text{and if } |\zeta_j| = 1, \text{ then it is a simple root}\}, \end{aligned} \quad (7)$$

with  $\overline{\mathbb{C}}$  denoting the extended complex plane.

We can regroup the terms in (4) as  $\Phi(\zeta, \mu) = \sum_{\ell=0}^k C_\ell(\mu) \zeta^\ell$  with some suitable polynomials  $C_\ell$ . The inequality condition in (3) implies that the leading coefficient  $C_k$  does not vanish identically; it may happen that for some exceptional  $\mu$  values the leading coefficient is zero:

$$\mathcal{E} := \{\mu \in \mathbb{C} : C_k(\mu) = 0\}.$$

For example, for the implicit Euler (IE) method  $\Phi(\zeta, \mu) = \varrho(\zeta) - \mu\sigma(\zeta) = (1 - \mu)\zeta - 1$  with  $\varrho(\zeta) := \zeta - 1$  and  $\sigma(\zeta) := \zeta$ , so  $\mathcal{E} = \{1\}$ . For the 2-step BDF method (BDF2),  $\Phi(\zeta, \mu) = (3 - 2\mu)\zeta^2 - 4\zeta + 1$ ; hence,  $\mathcal{E} = \{3/2\}$ . If definition (6) (or (7)) is interpreted formally, we have for the IE method that  $\mathcal{E} = \{1\} \subset \mathcal{S}$  (because (6) is satisfied vacuously). Similarly, for the BDF2 method,  $\mathcal{E} = \{3/2\} \subset \mathcal{S}$  (because then the unique root of  $\Phi(\zeta, 3/2) = 0$  is  $\zeta = 1/4$ ).

However, elements of  $\mathcal{E}$  or  $\mathcal{E} \cap \mathcal{S}$  can be problematic.

- (i) For  $\mu \in \mathcal{E}$ , the order of the recursion (3) decreases, thus, in general, the starting values  $y_0, y_1, \dots, y_{k-1}$  of the numerical method cannot be chosen arbitrarily.
- (ii) Some exceptional values  $\mu \in \mathcal{E} \cap \mathcal{S}$  can be surrounded by points of instability of the method—this is the case for example for both the IE and BDF2 methods. When the step size  $h > 0$  is chosen in a way that  $\mu \in \mathcal{E} \cap \mathcal{S}$  is such an isolated value, the recursion (3) generated by the numerical method becomes practically useless (it quickly “blows up” for arbitrarily small perturbations of  $h$ ).
- (iii) RLCs are often used to identify the boundary  $\partial\mathcal{S}$  of the stability region (see Sections 2.4–2.5 below). In [27, Definition (2.21)], the RLC is given by

$$\Gamma := \{\mu \in \overline{\mathbb{C}} : \exists \zeta \in \mathbb{C} \text{ with } |\zeta| = 1 \text{ and } \Phi(\zeta, \mu) = 0\}.$$

It can happen that  $\partial\mathcal{S}$  is a proper subset of the corresponding RLC (see, for example, our Fig. 3), but in [27, Corollary 2.6] it is shown that for a numerical method satisfying *Property C* (see [27, Formula (2.9)] or [17, Definition 4.7]), the RLC coincides with  $\partial\mathcal{S}$ . According to [17, Section V.4], all one-step methods have Property C, so the IE method also has. And indeed, applying [27, Proposition 2.7] to the IE method we now have that  $\varrho$  and  $\sigma$  have no common root and  $\varrho/\sigma$  is univalent on the set  $\{z \in \overline{\mathbb{C}} : |z - 1| > 1\}$ , so  $Q(\mu) = 1/(1 - \mu)$  has Property C. Thus for the IE method  $\partial\mathcal{S} = \Gamma$ . As we have seen above,  $1 \in \mathcal{E} \cap \mathcal{S}$ , so  $1 \in \partial\mathcal{S}$ . On the other hand,  $\Phi(\zeta, 1) = \varrho(\zeta) - \sigma(\zeta) = -1$ , so  $1 \notin \Gamma = \partial\mathcal{S}$ . This apparent contradiction seems to indicate that the authors of [27] interpreted definition (7) intuitively: a root  $\zeta = \infty$  is tacitly introduced as soon as the leading coefficient  $C_k(\mu)$  becomes zero. So [27, Corollary 2.6], for example, actually relies on definition (5) rather than on definition (7) (or (6)).

The problem of vanishing leading coefficient is implicitly avoided in [33, p. 66], or in [40], because they impose a requirement on “all the roots  $r_s$  ( $s = 1, \dots, k$ ).” Definition (5) above with the non-vanishing leading coefficient essentially appears, for example, in [41, Section 2.1] (where it is formulated for LMMs, that is, for  $s = 1$  in (3)), or in [42, Section 2].

Notice that, with the theorems cited in our Section 2.2, one can directly investigate the stability region of a numerical method, *without* constructing the corresponding RLC or without analyzing the relation between  $\partial\mathcal{S}$  and the RLC (see Sections 5–6 below).

Finally, we remark that the above considerations also play an important role, e.g., in control theory [2, Chapter 1], where a “degree invariance” (i.e., “no degree loss”) condition is incorporated in the *Boundary Crossing Theorem*. Bhattacharyya et al. [2, Chapter 1] also recalls several stability results for polynomials, e.g., the Routh–Hurwitz, Jury, or the recursive Schur(–Cohn) stability tests.

## 2.4 The RLC of a LMM

A linear multistep method for (2) has the form

$$\sum_{j=0}^k (\alpha_j y_{n+j} - h\beta_j f_{n+j}) = 0, \quad (8)$$

where  $f_m := f(t_m, y_m)$ , and the numbers  $\alpha_j \in \mathbb{R}$  and  $\beta_j \in \mathbb{R}$  ( $j = 0, \dots, k$ ) are the suitably chosen method coefficients with  $\alpha_k \neq 0$ . The method is implicit, if  $\beta_k \neq 0$ . By setting

$$\varrho(\zeta) := \sum_{j=0}^k \alpha_j \zeta^j \quad \text{and} \quad \sigma(\zeta) := \sum_{j=0}^k \beta_j \zeta^j,$$

the associated characteristic polynomial (4) becomes

$$\Phi(\zeta, \mu) \equiv P_1(\zeta, \mu) := \varrho(\zeta) - \mu\sigma(\zeta). \quad (9)$$

One way to study the stability region (5), or its boundary  $\partial\mathcal{S}$  in the complex plane is to depict the RLC corresponding to the method [17]: observe that  $P_1$  is linear in  $\mu$ , so  $P_1(\zeta, \mu) = 0$  implies  $\mu = \varrho(\zeta)/\sigma(\zeta)$  (for  $\sigma(\zeta) \neq 0$ ). The RLC is then the image of the parametric curve

$$[0, 2\pi] \ni \vartheta \mapsto \mu(\vartheta) := \frac{\varrho(e^{i\vartheta})}{\sigma(e^{i\vartheta})}. \quad (10)$$

### 2.4.1 RLCs for the BDF methods

Each member of the BDF family is a special case of (8). The  $k$ -step BDF method (having order  $k$ ) is given by

$$\sum_{j=1}^k \frac{1}{j} \nabla^j y_{n+1} = h f_{n+1},$$

where  $\nabla$  denotes the backward difference operator  $\nabla y_{n+1} := y_{n+1} - y_n$ , and  $\nabla^j y_{n+1} := \nabla^{j-1} y_{n+1} - \nabla^{j-1} y_n$  (for  $j > 1$ ). It is known [17] that the corresponding RLC is

$$\mu(\vartheta) \equiv \sum_{j=1}^k \frac{1}{j} (1 - e^{-i\vartheta})^j. \quad (11)$$

Figures 1, 2, and 3 show the RLCs for some BDF methods.

### 2.5 The RLC of a multiderivative multistep method

A second-derivative multistep method is more general than (8) and can be written as

$$\sum_{j=0}^k (\alpha_j y_{n+j} - h\beta_j f_{n+j} - h^2\gamma_j g_{n+j}) = 0, \quad (12)$$

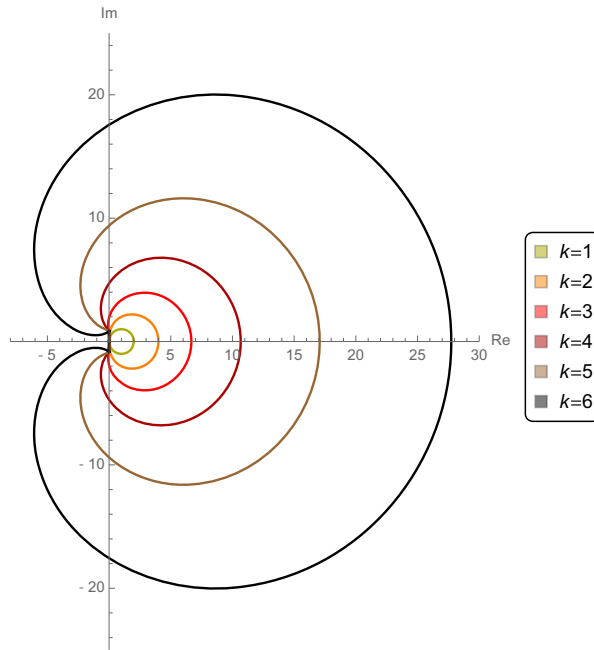
where  $g_n := g(t_n, y_n)$  with  $g(t, y) := \partial_1 f(t, y) + \partial_2 f(t, y) \cdot f(t, y)$ , and the method is determined by the coefficients  $\alpha_j$ ,  $\beta_j$ , and  $\gamma_j$ , see [17]. Now the associated characteristic polynomial (4) becomes

$$\Phi(\zeta, \mu) \equiv P_2(\zeta, \mu) := \sum_{j=0}^k (\alpha_j - \mu\beta_j - \mu^2\gamma_j)\zeta^j.$$

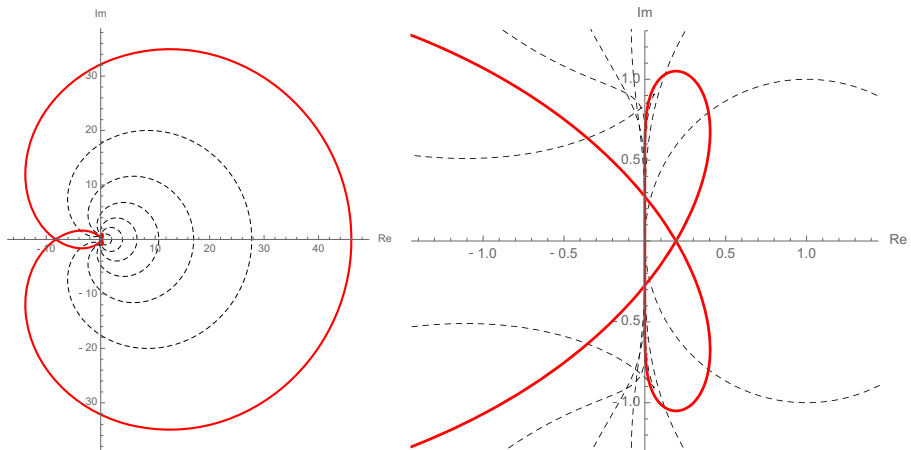
This time we have two RLCs:

$$[0, 2\pi] \ni \vartheta \mapsto \mu_{1,2}(\vartheta), \quad (13)$$

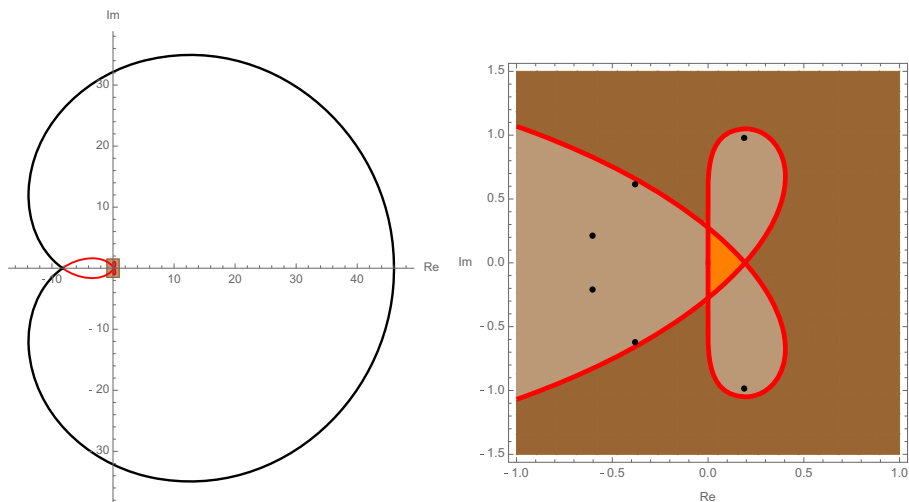
where  $\mu_{1,2}$  are the two solutions of  $P_2(e^{i\vartheta}, \mu) = 0$ . For any choice of the method coefficients  $\alpha_j$ ,  $\beta_j$ , and  $\gamma_j$ , one can construct  $\mu_{1,2}$  explicitly, since  $P_2$  is only quadratic in  $\mu$ .



**Fig. 1** RLCs for the  $k$ -step BDF methods for  $1 \leq k \leq 6$ . The stability region of the method in each case is the unbounded component of  $\mathbb{C}$



**Fig. 2** RLC for the unstable 7-step BDF method in red (left) and a close-up near the origin (right). For comparison, the curves from Fig. 1 are also superimposed as dashed gray curves



**Fig. 3** The black curve in the left figure shows the boundary  $\partial S$  of the stability region of the (unstable) 7-step BDF method;  $\partial S$  is non-differentiable at one point. The stability region is the unbounded outer component. The red curve segment near the origin is not part of  $\partial S$ ; it is a subset only of the RLC as displayed in Fig. 2. The small brown rectangle in the center is shown in detail in the right figure. The red curve in the right figure is again the RLC. The 6 black dots depict the set of  $\mu$  values such that  $P_1(\cdot, \mu)$  in (9) has multiple roots (there are no other  $\mu \in \mathbb{C}$  parameters with this property for  $k = 7$ ). The polynomial  $P_1(\cdot, \mu)$  has 1, 2, and 3 roots outside the unit disk for  $\mu$  values in the dark brown, light brown and orange regions, respectively;  $P_1(\cdot, \mu)$  cannot have 4 or more roots outside the unit disk. Each of the three self-intersections of the RLC in this figure (as well as the self-intersection of the RLC seen only in the left figure) corresponds to a  $\mu$  value for which  $P_1(\cdot, \mu)$  has two distinct roots with modulus 1. Exactly computing, for example, the unique value of  $\mu_{\dagger} \approx -2.68886 \cdot 10^{-6} + 0.275988i$  in the open upper half-plane where the RLC crosses itself was a non-trivial task: it took *Mathematica* 86 minutes to explicitly determine the coefficients of the integer polynomial defining  $\mu_{\dagger}$  and having degree 30. The RLCs for the  $k$ -step BDF methods with  $1 \leq k \leq 6$  do not have any self-intersections; other singularities may occur, see Fig. 6

### 2.5.1 RLCs for the Enright methods

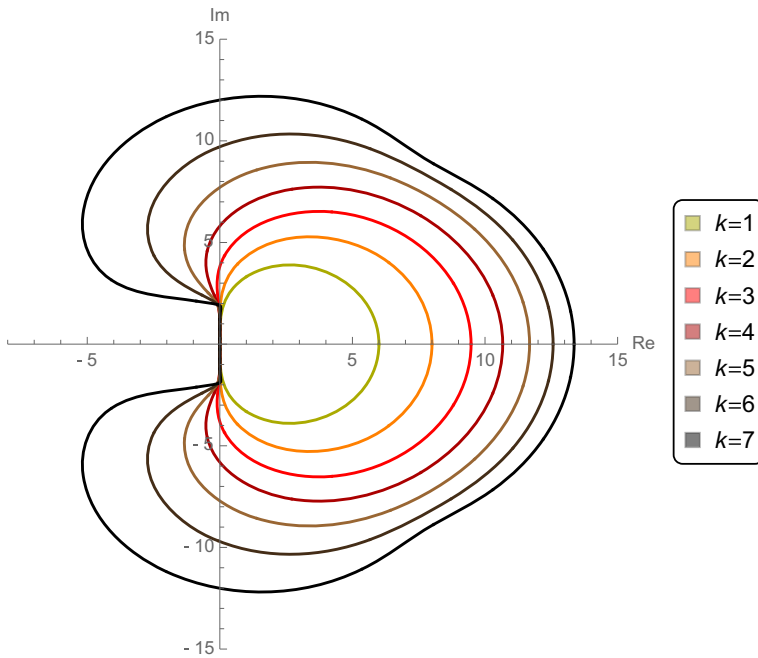
The Enright methods are special cases of (12), and for  $k \geq 1$  they are defined [17] as

$$y_{n+1} = y_n + hf_{n+1} - h \sum_{j=1}^k \left( \frac{1}{j} \left( \sum_{\ell=j}^k v_{\ell} \right) \nabla^j f_{n+1} \right) + h^2 \left( \sum_{\ell=0}^k v_{\ell} \right) g_{n+1}, \quad (14)$$

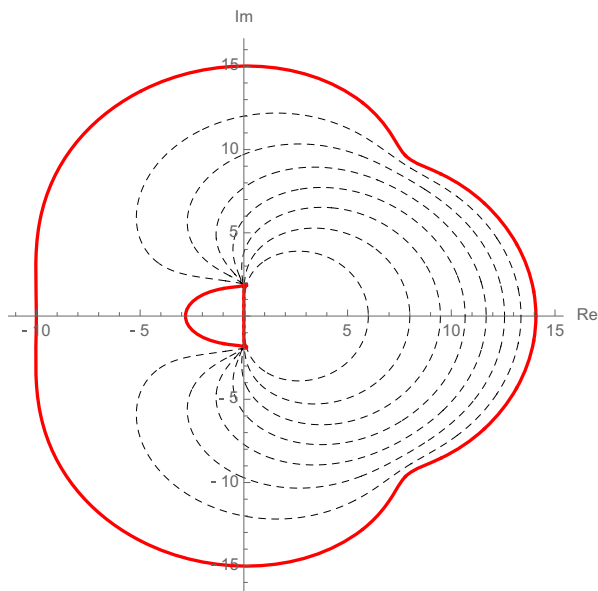
where

$$v_{\ell} := (-1)^{\ell} \int_0^1 (\tau - 1) \binom{1-\tau}{\ell} d\tau \quad (0 \leq \ell \leq k)$$

with the usual extension of the binomial coefficients. From (14) one obtains the RLCs of the Enright methods, see Figs. 4 and 5. The order of the  $k$ -step Enright method is  $k + 2$ .



**Fig. 4** RLCs for the  $k$ -step Enright methods for  $1 \leq k \leq 7$ . The stability region of the method in each case is the unbounded component of  $\mathbb{C}$



**Fig. 5** RLCs for the unstable 8-step Enright method in red. The stability region  $\mathcal{S}$  is not connected,  $\mathbb{C} \setminus \mathcal{S}$  is the annulus-like region. For comparison, the curves from Fig. 4 are displayed as dashed gray curves

### 3 Optimal sector inclusions

#### 3.1 The RLC in implicit algebraic form

Computing the stability angle of a method with stability region  $\mathcal{S}$  is equivalent to finding the slope of the unique line  $L$  that passes through the origin, touches  $\partial\mathcal{S}$  at some point in the open upper left half-plane such that  $\partial\mathcal{S}$  lies on the right-hand side of  $L$  (viewed from the origin) in this quadrant. This last requirement is necessary since  $\partial\mathcal{S} \cap L$  can consist of more points, even in the open upper left half-plane (see Fig. 7).

Assume now that  $\partial\mathcal{S}$  can be represented by the RLC of the method (cf. Remark 2.7). As we have seen, the RLC is the image of the function  $\mu(\cdot)$  in (10) for LMMs, or the union of the images of the functions  $\mu_{1,2}(\cdot)$  in (13) for second-derivative multistep methods. The function  $\mu$  is given as a simple ratio, but to get the explicit forms of  $\mu_{1,2}$ , one should solve a quadratic equation. As the value of  $k$  gets larger, these explicit formulae for  $\mu_{1,2}$  corresponding to a  $k$ -step second-derivative multistep method become more and more complicated. Moreover, obtaining explicit and practically useful parametrized formulae for the RLCs associated with multistep methods based on higher-than-second-order derivatives would be almost impossible.

To avoid these difficulties, we now describe a more general and effective technique which reduces the determination of the stability angles to the solution of a suitable system of polynomial equations. Let us consider the equation  $\Phi(e^{i\vartheta}, \mu) = 0$  (see (4)). By using the well-known Weierstrass substitution [43, pp. 382–383]

$$\vartheta = 2 \arctan(t) \quad (t \in \mathbb{R}),$$

we have  $e^{i\vartheta} = (i - t)/(i + t)$ ; so instead of solving  $\Phi(e^{i\vartheta}, \mu) = 0$  for  $\mu$ , we can solve

$$\Phi\left(\frac{i - t}{i + t}, \mu\right) = 0 \quad (15)$$

without trigonometric functions. Notice that originally we have  $\vartheta \in [0, 2\pi]$  in  $e^{i\vartheta}$ , or equivalently,  $\vartheta \in (-\pi, \pi]$ , but  $\pi$  is not in the range of the function  $2 \arctan$ ; therefore, we define

$$M_{-1} := \{\mu \in \mathbb{C} : \Phi(e^{i\pi}, \mu) = 0\}$$

to restore the missing  $\mu$  value(s) due to the reparametrization. Then, clearly, (15) can be brought to the form  $Q(t, \mu)/R(t) = 0$  with some (complex) polynomials  $Q$  and  $R$ . By writing  $\mu = a + bi$  ( $a, b \in \mathbb{R}$ ), we get that there exist two real polynomials  $Q_{\text{re}} : \mathbb{R}^3 \rightarrow \mathbb{R}$  and  $Q_{\text{im}} : \mathbb{R}^3 \rightarrow \mathbb{R}$  such that the solutions of  $Q(t, \mu) = 0$  for any fixed  $t \in \mathbb{R}$  are obtained as the solutions of the system

$$\begin{cases} Q_{\text{re}}(t, a, b) = 0 \\ Q_{\text{im}}(t, a, b) = 0. \end{cases} \quad (16)$$

Now, we eliminate  $t$  by taking the resultant [14] of  $Q_{\text{re}}$  and  $Q_{\text{im}}$  with respect to this parameter, and get that there exists a real polynomial  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  such that if (16) holds for some  $t \in \mathbb{R}$ , then  $F(a, b) = 0$  should hold with some  $a, b \in \mathbb{R}$ . Hence, after identifying  $\mathbb{C}$  with  $\mathbb{R}^2$ , we see that the RLC can be represented as the implicit algebraic curve  $C \cup M_{-1}$  with  $C := \{(a, b) \in \mathbb{R}^2 : F(a, b) = 0\}$ . Assuming

that the set  $M_{-1}$  is finite (it has at most two elements in the case of the BDF and Enright methods we are interested in), we ignore this component and focus only on  $C$ . Suppose now that a line  $L$  passes through the origin and touches  $C$  in the open upper left half-plane at some  $(a_0, b_0)$  with  $a_0 < 0 < b_0$ . By assuming that  $C$  can be represented locally as the graph of an implicit function near  $(a_0, b_0) \in C$ , we easily get, by differentiating  $a \mapsto F(a, b(a))$ , that  $(a_0, b_0)$  satisfies

$$\begin{cases} F(a_0, b_0) &= 0 \\ a_0 \cdot \partial_1 F(a_0, b_0) + b_0 \cdot \partial_2 F(a_0, b_0) &= 0 \\ a_0 &< 0 \\ b_0 &> 0. \end{cases} \quad (17)$$

By taking again the resultant of the first two polynomial equations, one of the variables, say  $b_0$ , is eliminated. The resulting univariate polynomial yields in the general case finitely many possible  $a_0$  values to choose from. With  $\alpha$  denoting the angle (in radians) between  $L$  and the negative half of the real axis, we get that  $\tan(\alpha) = -b_0/a_0$ . To select the appropriate solution  $(a_0, b_0)$  (and hence the appropriate tangent line  $L$ ), we verify in the concrete case that  $(a_0, b_0) \in \partial S \subset \mathbb{C} = \mathbb{R}^2$ , and determine whether  $\partial S$  lies on the right-hand side of  $L$ . The appropriately chosen  $\alpha$  angle then yields the desired stability angle.

### 3.2 Results for the BDF methods

The simplest non-trivial case illustrating the steps in Section 3.1 is the determination of the stability angle for the 3-step BDF method. Formula (11) with  $k = 3$  yields the following trigonometric parametrization of the RLC in  $\mathbb{R}^2$  after a simplification:

$$[0, 2\pi] \ni \vartheta \mapsto \mu(\vartheta) := \left( \frac{4}{3} \sin^4\left(\frac{\vartheta}{2}\right) (1 - 4 \cos(\vartheta)), \frac{\sin(\vartheta)}{3} [2(\cos(2\vartheta) + 5) - 9 \cos(\vartheta)] \right).$$

After eliminating the trigonometric functions, (15) can be written as

$$\frac{Q(t, \mu)}{R(t)} = \frac{3\mu t^3 - 20t^3 - 9\mu t + 6t + i(3\mu - 9\mu t^2 + 18t^2)}{3(t - i)^3} = 0.$$

Then  $Q_{\text{re}}$  and  $Q_{\text{im}}$  in (16) become

$$\begin{cases} 3at^3 - 9at + 9bt^2 - 3b - 20t^3 + 6t = 0 \\ -9at^2 + 3a + 3bt^3 - 9bt + 18t^2 = 0. \end{cases}$$

We eliminate  $t$  from this system and obtain

$$F(a, b) := 432 \left[ 108a^6 - 1188a^5 + 9a^4 (36b^2 + 439) - 2a^3 (1188b^2 + 3121) + 9a^2 (36b^4 + 394b^2 + 547) - 54a (22b^4 + 17b^2 + 30) + 27b^4 (4b^2 - 15) \right].$$

Now,  $b$  is eliminated from the first two equations of (17), and we get that the possible choices for  $a_0$  are the negative real roots of

$$a^4(24a - 25)^4(5324a + 405)^2(6a^2 - 13a + 9)^2 = 0,$$



yielding the unique value  $a_0 = -405/5324$ . Substituting this  $a_0$  into (17), we get the unique value  $b_0 = 987\sqrt{35}/5324$ ; hence,  $\tan(\alpha) = -b_0/a_0 = (329\sqrt{7/5})/27$  is the only possible value for the tangent of the stability angle. Finally, we verify that the corresponding tangent line  $L$  passing through the origin has no other intersection point with  $\partial\mathcal{S}$  in the open upper left quadrant, and  $\partial\mathcal{S}$  lies on the right side of  $L$ .

**Remark 3.1** The above RLC for the 3-step BDF method can also be parametrized as

$$\mathbb{R} \ni t \mapsto \left( \frac{4t^4(5t^2 - 3)}{3(t^2 + 1)^3}, \frac{2t(21t^4 + 8t^2 + 3)}{3(t^2 + 1)^3} \right) \in \mathbb{R}^2.$$

Here,  $M_{-1} = \{(20/3, 0)\} \subset \mathbb{R}^2$ , corresponding to the  $t \rightarrow \pm\infty$  limiting value of the parametrization.

The remaining stability angle values for  $4 \leq k \leq 6$  can be computed analogously, so Table 1 shows only the final exact results.

**Remark 3.2** For  $3 \leq k \leq 6$ , the BDF stability region includes an interval along the imaginary axis and containing the origin if and only if  $k = 5$  or  $k = 6$ . For  $k = 5$  and  $k = 6$  the two intervals are

$$\{z \in \mathbb{C} : \operatorname{Re}(z) = 0, |\operatorname{Im}(z)| \leq \frac{1}{12\sqrt{2}}\sqrt{12775 - 387\sqrt{1065}} \approx 0.710\} \subset \mathcal{S}$$

and

$$\{z \in \mathbb{C} : \operatorname{Re}(z) = 0, |\operatorname{Im}(z)| \leq \frac{7}{20}\sqrt{1263 - 336\sqrt{14}} \approx 0.843\} \subset \mathcal{S},$$

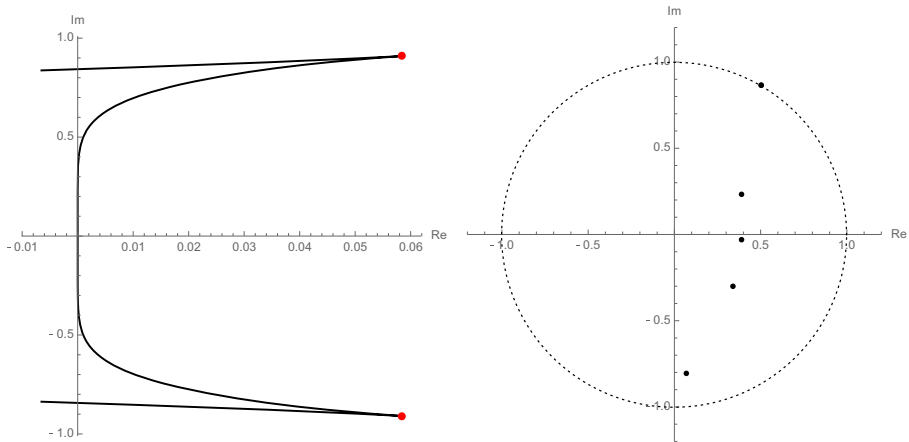
respectively.

**Remark 3.3** The boundary curve of the stability region of the 6-step BDF method contains two cusp singularities (see Fig. 6 and compare with Fig. 3). No other  $\partial\mathcal{S}$  curve has this type of degeneracy in the BDF family for  $1 \leq k \leq 5$  or  $k = 7$ . Since the cusp points for  $k = 6$  are not part of  $\mathcal{S}$ , the stability region in this case is not closed (nor open).

### 3.3 Results for the Enright methods

By applying the algorithm described in Section 3.1, we can exactly determine the stability angles for the Enright methods (see Table 2). But since the  $c_k^{\text{Enr}}$  values are much more complicated algebraic numbers than the corresponding  $c_k^{\text{BDF}}$  constants in Table 1, Table 2 contains only a numerical approximation to the exact stability angles.

**Remark 3.4** By rounding the values of  $\alpha_k^{\text{Enr}}$  given in Table 2 to two decimal places, we recover the approximate values of these stability angles in [17, Chapter V.3, Table 3.1].



**Fig. 6** Cusp singularities of  $\partial S$  for the 6-step BDF method, denoted by red dots in the left figure. The singularities are located at  $\mu_{\pm} := \frac{7}{120} \pm i \frac{21\sqrt{3}}{40} \approx 0.0583 \pm 0.9093i$ . For each such  $\mu$  value,  $P_1(\cdot, \mu)$  in (9) has a double root with modulus equal to 1. Therefore,  $\mu_{\pm} \in \partial S \setminus S$ ; hence, this  $S$  is not closed. The right figure depicts the 6 roots of  $P_1(\cdot, \mu_+)$ , and the double root is located at  $\frac{1}{2}(1 + i\sqrt{3})$  (note that  $\mu_+ \in \mathbb{C} \setminus \mathbb{R}$ , so these roots are not symmetric with respect to the real axis)

It turns out that  $c_3^{\text{Enr}}$  is an algebraic number of degree 22, being the unique positive root of the following even polynomial with coefficients

$$\{662162550162672001197071902273445952000000000000000, 0, \\ 4744945665370497147850526235135397935643117766707200000, 0, \\ 74537179754361052063480563770102869789636567887828480000, 0, \\ 417809113212221868517393954677075422852686053100794277975, 0, \\ 1103592881533264097533512931940128409045933472020943607320, 0, \\ 1780216754145335084531442707748395556646595339402356863603, 0, \\ 2028417751642933570985301304414377204911584843581604760752, 0, \\ 1720629215811045658880293770988465046952673868659037700813, 0, \\ 1065257770963658030926145190690110109450795207237154063632, 0, \\ 451976742777053443392779380035051991794204051855298481913, 0, \\ 117280744006618927204325767614876515512652225395198902600, 0, \\ 14037302894263476230042573549418427869442188056651130000\}.$$

**Remark 3.5** Besides the stability angle, there are other measures of stability for  $A(\alpha)$ -stable methods. One of these characteristics is the stiff stability abscissa, being the smallest constant  $D > 0$  such that  $\{z \in \mathbb{C} : \text{Re}(z) \leq -D\} \subset S$ . For example, for the 3-step Enright method, Table 3.1 in [17, Chapter V.3] contains the approximate value  $D \approx 0.103$ . By using our implicit representation of  $\partial S$ , it is straightforward to determine the exact value of  $D \approx 0.10341810907195$ ; it is an algebraic number

of degree 12, and the total number of digits in the coefficients of its defining integer polynomial is 529.

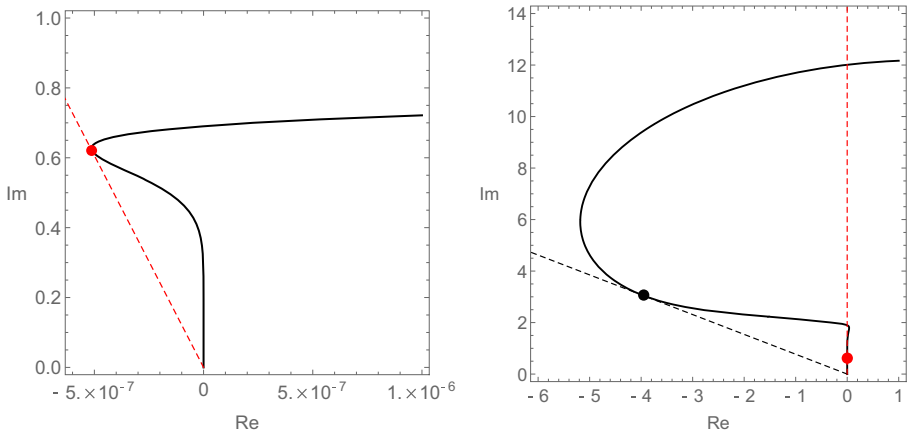
As for the  $k = 4$  case, the algebraic degree of  $c_4^{\text{Enr}}$  is 28. The constants  $c_5^{\text{Enr}}$ ,  $c_6^{\text{Enr}}$ , and  $c_7^{\text{Enr}}$  can be given as roots of increasingly more involved integer polynomials, so we do not reproduce these polynomials here. During the computations in the  $k = 7$  case, for example, we had to manipulate intermediate polynomials of degree of a few hundred, or polynomials with a total number of coefficient digits of approximately 470000. We could describe the final defining polynomial for  $c_7^{\text{Enr}}$  by  $\approx 175000$  characters in *Mathematica*.

**Remark 3.6** Let us consider the Enright stability region corresponding to  $k = 7$ . As we already remarked earlier, there are exactly two lines that pass through the origin and are locally tangent to the boundary curve at some point in the open upper left half-plane (see Fig. 7). Within the BDF family for  $1 \leq k \leq 6$  or in the Enright family for  $1 \leq k \leq 7$ , this phenomenon occurs only in the present case.

## 4 Optimal disk inclusions

As for the largest inscribed disk in the stability region  $\mathcal{S}$ , we again expect—similarly to Section 3.1—that  $\partial\mathcal{S}$  (or the RLC) and the optimal disk possess a common tangent line (with point of tangency different from the origin). By using:

- the implicit algebraic form  $F(a, b) = 0$  of the RLC



**Fig. 7** Part of the boundary of the stability region of the 7-step Enright method near the origin (solid black curve) together with the two (dashed red and black) lines that pass through the origin and are locally tangent to the boundary curve at some point in the open upper left half-plane. Due to the scaling, the dashed black line is seen only in the larger plot window on the right. The stability angle  $\alpha_7^{\text{Enr}} \approx 37.6^\circ$  of the method is determined by the dashed black line; the red line has additional intersection points with the boundary curve. The angle between the dashed red line and the negative half of the real axis has also been computed exactly; its approximate value is  $\approx 89.9999527^\circ$

- the implicit equation  $(a+r)^2 + b^2 - r^2 = 0$  for the boundary of the inscribed disk
- and the condition for a common tangent line

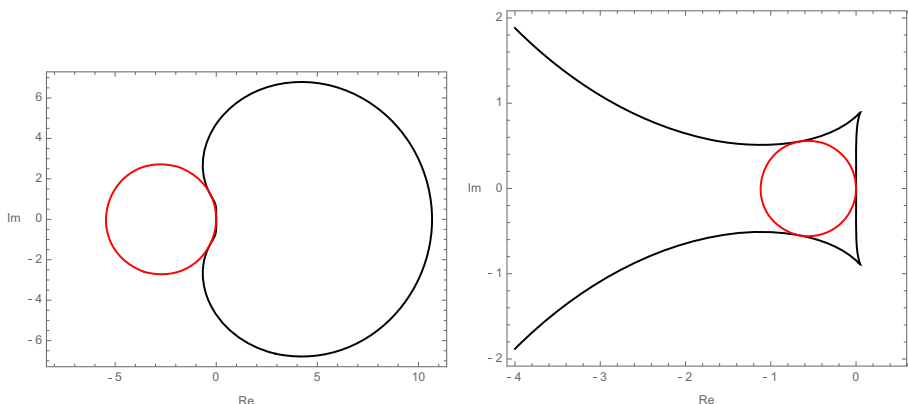
$$-\frac{\partial_a F(a, b)}{\partial_b F(a, b)} = -\frac{\partial_a ((a+r)^2 + b^2 - r^2)}{\partial_b ((a+r)^2 + b^2 - r^2)},$$

we obtain a system of 3 polynomial equations in 3 unknowns  $(a, b, r)$ . By taking resultants and successively eliminating the variables  $(a, b)$ , we obtain a univariate polynomial in  $r$  whose positive root will yield the optimum value of the stability radius. The exact optimal stability radii  $r_k^{\text{BDF}}$  for the  $k$ -step BDF methods ( $3 \leq k \leq 6$ ) are found in Table 3 (see also Fig. 8). The degree of the algebraic number  $r_k^{\text{BDF}}$  is 2, 3, 5, and 5 for  $3 \leq k \leq 6$ , respectively.

**Remark 4.1** It is quite surprising that the algebraic numbers listed in Table 3 have such a low degree for the following reasons. For the 3-step BDF method, the univariate polynomial in  $r$  mentioned above has degree 28, but it can be split into several factors of lower degree and has a unique positive root  $r_3^{\text{BDF}} \approx 7.0497$ . For the 4-step BDF method, the corresponding  $r$ -polynomial has degree 52 and a unique positive root  $\approx 2.7272$ . The  $r$ -polynomial for the 5-step BDF method has degree 88 and a unique positive root  $\approx 1.3579$ . Finally, the  $r$ -polynomial for the 6-step BDF method has degree 128 and a unique positive root  $\approx 0.5599$ .

## 5 Optimal stability angle in a family of multistep methods

In [18], ODEs of the form  $u'(t) = F(u(t)) + G(u(t))$ ,  $u(0) = u_0$  are considered, with  $F$  and  $G$  representing non-stiff and stiff parts of the equation, respectively. To solve these equations numerically, the authors construct several implicit-explicit (IMEX)



**Fig. 8** The largest inscribed disk  $|z+r| \leq r$  (with red boundary) in the stability region of the  $k$ -step BDF method for  $k=4$  (left) and  $k=6$  (right), see Table 3

LMMs and thoroughly analyze them from the viewpoint of numerical monotonicity, boundedness, and stability. Their analysis involves finding optimal methods with respect to various criteria in certain families.

Here, we take their simplest case study from [18, Section 3.2.1], a 2nd-order, 3-step explicit method augmented by an implicit method (note that the time step is now denoted by  $\Delta t$  instead of  $h$ , and we changed their notation from  $b_j$  to  $\beta_j$ ):

$$u_n = \frac{3}{4}u_{n-1} + \frac{1}{4}u_{n-3} + \frac{3}{2}\Delta t \cdot F_{n-1} + \sum_{j=0}^3 \beta_j \Delta t \cdot G_{n-j}. \quad (18)$$

The values of  $\beta_2 := -3\beta_0 - 2\beta_1 + 3$  and  $\beta_3 := 2\beta_0 + \beta_1 - \frac{3}{2}$  are determined from the order conditions, so (18) becomes a 2-parameter family of methods, with real parameters  $\beta_1$  and  $\beta_0$ . The three figures in [18, Figure 1] then depict the  $A(\alpha)$ -stability angles, the “damping factors,” and the “absolute error constants,” respectively, of members of the family (18). In what follows, we do not consider these last two categories but focus only on the leftmost figure in [18, Figure 1]—as the authors conclude in [18, Section 3.2.1], a method with large stability angle does not necessarily have a good damping factor or a small error constant, and vice versa; the different optimization criteria are often conflicting. In other words, our goal in this section is to find the IMEX method in the family (18) with the largest stability angle.

To begin the  $A(\alpha)$ -stability investigation, the authors of [18] define the usual linear test functions  $F(u) := \hat{\lambda}u$  and  $G(u) := \lambda u$ . They then assume that  $\Delta t \cdot \hat{\lambda} = i\eta$  and  $\Delta t \cdot \lambda = \xi$  with  $\eta \in \mathbb{R}$  and  $\mathbb{R} \ni \xi \leq 0$ : this choice is relevant “for example, for advection-diffusion equations if central finite differences or spectral approximations are used in space.” These assumptions lead to the following characteristic polynomial of the IMEX multistep family (see [18, (2.4)–(2.7)]):

$$\mathbb{C} \ni \zeta \mapsto \zeta^3 - \left(\frac{3}{4}\zeta^2 + \frac{1}{4}\right) - i\eta \left(\frac{3}{2}\zeta^2\right) - \xi \left(\sum_{j=0}^3 \beta_j \zeta^{3-j}\right). \quad (19)$$

To create the leftmost figure in [18, Figure 1] approximately indicating the optimal stability angle within the family, the authors use (19) to construct the RLCs and study these curves “for  $\xi \rightarrow -\infty$ ” to estimate the stability angles.<sup>1</sup>

In the rest of this section, we confirm their numerical findings, but we solve the optimization problem rigorously and exactly. We have selected family (18) because the final result—the optimal stability angle—has a particularly simple form (see our Theorem 5.3 below), and, at the same time, our straightforward approach based on the theorems cited in Section 2.2 is readily illustrated. We emphasize that our analysis avoids the construction of the RLCs: as we have seen (for example, in Fig. 3), they

<sup>1</sup>When the stability angle  $\alpha$  of a method is defined in [18, Sections 2.3 and 3.2.1] notice that we should require that the sector

$$\xi \leq 0, \quad |\eta/\xi| \leq \tan(\alpha) \quad \text{with angle } \alpha \leq \pi/2$$

be included in the stability region in the  $(\xi, \eta)$ -plane (with the  $\xi = 0$  and  $\alpha = \pi/2$  cases interpreted appropriately). In other words,  $\arctan(\alpha)$  in [18] is to be replaced by  $\tan(\alpha)$ ; otherwise, the sector would not “open wide enough” and  $A$ -stability would not be recovered in the  $\alpha \rightarrow \pi/2^-$  limit. See also Footnote 2.

may have complicated self-intersections, and it is often not obvious a priori whether a particular segment of the RLC coincides with the stability region boundary or not.

### 5.1 Summary of the main steps and results

By rearranging (19) and inserting the values of  $\beta_2$  and  $\beta_3$  given below (18), we define

$$P_{\beta_1, \beta_0}(\zeta, \xi, \eta) := (1 - \beta_0 \xi) \zeta^3 - \left( \frac{3}{4} + \beta_1 \xi + \frac{3i\eta}{2} \right) \zeta^2 + \xi (3\beta_0 + 2\beta_1 - 3) \zeta - \left( \frac{1}{4} + 2\beta_0 \xi + \beta_1 \xi - \frac{3}{2} \xi \right), \quad (20)$$

where  $\zeta \in \mathbb{C}$ ,  $(\beta_1, \beta_0) \in \mathbb{R}^2$ ,  $\xi \leq 0$  and  $\eta \in \mathbb{R}$ . Our goal is to find the parameters  $(\beta_1, \beta_0)$  such that the stability region

$$\mathcal{S}_{\beta_1, \beta_0} := \{(\xi, \eta) \in \mathbb{R}^2 : \xi \leq 0, \eta \in \mathbb{R}, P_{\beta_1, \beta_0}(\cdot, \xi, \eta) \in \mathbf{svN}\} \quad (21)$$

contains the infinite sector

$$\mathcal{A}_m := \{(\xi, \eta) \in \mathbb{R}^2 : \xi \leq 0, \eta \in \mathbb{R}, |\eta| \leq m|\xi|\}$$

with the largest  $m > 0$  in the definition of  $A(\alpha)$ -stability. In other words, we are to find  $(\beta_1, \beta_0)$  such that

$$\mathcal{A}_m \subset \mathcal{S}_{\beta_1, \beta_0} \quad (22)$$

holds with the largest possible  $m > 0$ . Note that for convenience we have identified  $\mathbb{C}$  with  $\mathbb{R}^2$ ; hence, stability regions in this section are subsets of  $\mathbb{R}^2$ .

As a first step, Lemma 5.1 below yields a necessary condition for the inclusion (22). In its proof—presented in Appendix A.1—we use the argument proposed in [18] and consider the  $\xi \rightarrow -\infty$ ,  $\eta = 0$  limiting values. At this point, it is convenient to recall the notion of  $\overset{\circ}{A}$ -stability [17, Chapter V.2]: a method is  $\overset{\circ}{A}$ -stable, if its stability region includes the non-positive reals  $\{\xi \in \mathbb{R} : \xi \leq 0\}$ . Clearly,

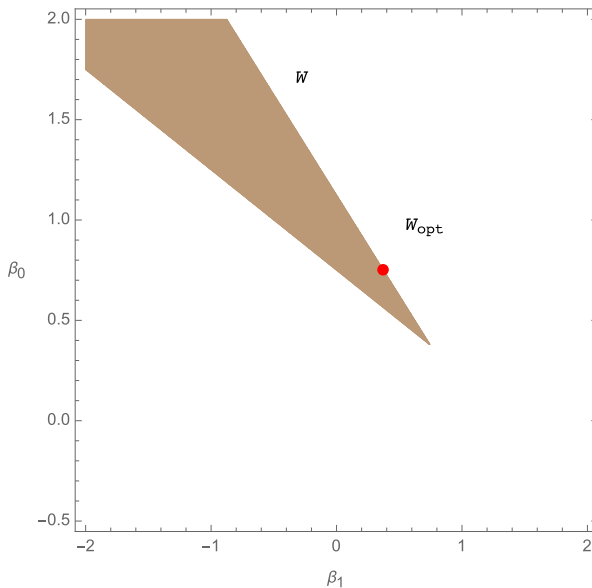
$$A(\alpha)\text{-stability with some } \alpha > 0 \implies \overset{\circ}{A}\text{-stability.}$$

**Lemma 5.1** *Let us define*

$$W := \left\{ (\beta_1, \beta_0) : \beta_1 \leq \frac{3}{4}, \frac{3 - 2\beta_1}{4} \leq \beta_0 \leq \frac{9 - 8\beta_1}{8} \right\}. \quad (23)$$

*Then a method of the form (18) is not  $\overset{\circ}{A}$ -stable for  $(\beta_1, \beta_0) \notin W$ .*

As a consequence, from now on, we can assume  $(\beta_1, \beta_0) \in W$  (see Fig. 9). Note that the orientation of the axes in Fig. 9 and in the leftmost figure in [18, Figure 1] is the same: the  $\beta_1$ -axis is horizontal, while the  $\beta_0$ -axis is vertical. Lemma 5.1 thus also proves that the wedge-like object in the parameter space in the leftmost figure in [18, Figure 1] is indeed a perfect (infinite) wedge given by  $W$ .



**Fig. 9** The figure shows the set  $W$  defined in (23) and the optimal parameter choice  $W_{\text{opt}}$  with  $(\beta_1, \beta_0) = (3/8, 3/4)$  determined in Appendix A.2. Interestingly, the coordinates of the vertex of the wedge  $W$  are  $(3/4, 3/8)$

**Remark 5.2** The assumption  $(\beta_1, \beta_0) \in W$  implies  $\beta_0 > 0$ , so due to  $\xi \leq 0$ , the leading coefficient of (20),  $1 - \beta_0 \xi$ , cannot vanish (cf. Remark 2.7).

Then in Appendix A.2 we prove the main result of Section 5.

**Theorem 5.3** Suppose that  $(\beta_1, \beta_0) \in W$ . Then the largest  $m > 0$  such that (22) holds is  $m \equiv m_{\text{opt}} := 1/2$ .

In the proof, we show that finding the optimal  $(\beta_1, \beta_0) \in W$  is equivalent to finding the largest positive real root of a suitable polynomial in  $m$  with coefficients depending on  $\beta_1$  and  $\beta_0$ . We verify that this optimal root is located at  $m_{\text{opt}}$ , corresponding to the unique method with  $(\beta_1, \beta_0) = W_{\text{opt}} := (3/8, 3/4) \in W$  and represented as a red dot in the parameter space in Fig. 9. The black curve in the left half-plane in Fig. 12 is the boundary of the optimal stability region, and the dashed red lines bound the largest inscribed infinite sector  $\mathcal{A}_{1/2}$ : the optimal stability angle satisfies  $\tan(\alpha) = m_{\text{opt}}$ . As a conclusion, the highest value in the scale adjacent to the leftmost figure in [18, Figure 1] should be exactly  $\alpha = \arctan(1/2) \approx 0.463648$ , that is,  $\alpha \approx 26.5651^\circ$ .

**Remark 5.4** Unlike in Section 6 (see Remark B.2), the boundary of the optimal sector  $\mathcal{A}_{1/2}$  does not touch (or intersect) the boundary of the optimal stability region  $\mathcal{S}_{3/8, 3/4}$  in the open left half-plane.

**Remark 5.5** In [18, Section 3.2.1, (3.4)–(3.5)], the stability angles for two particular schemes from the family (18) are also approximated. For the IMEX-Shu(3,2) scheme

$$u_n = \frac{3}{4}u_{n-1} + \frac{1}{4}u_{n-3} + \frac{3}{2}\Delta t \cdot F_{n-1} + \frac{4}{9}\Delta t \cdot G_n + \frac{2}{3}\Delta t \cdot G_{n-1} + \frac{1}{3}\Delta t \cdot G_{n-2} + \frac{1}{18}\Delta t \cdot G_{n-3}$$

they obtain  $\alpha_{\text{Shu}} \approx 0.06$ , and for the IMEX-SG(3,2) scheme

$$u_n = \frac{3}{4}u_{n-1} + \frac{1}{4}u_{n-3} + \frac{3}{2}\Delta t \cdot F_{n-1} + \Delta t \cdot G_n + \frac{1}{2}\Delta t \cdot G_{n-3}$$

they get  $\alpha_{\text{SG}} \approx 0.38$ . Our technique easily yields the exact values

$$\alpha_{\text{Shu}} = \arctan\left(1/\sqrt{135 + 78\sqrt{3}}\right) \approx 0.0607719,$$

and

$$\alpha_{\text{SG}} = \arctan\sqrt{\frac{1}{3}(2\sqrt{3} - 3)} \approx 0.374734.$$

## 6 Optimal parabola inclusion in a family of multistep methods

In the previous section, we demonstrated how one can find the optimal sector in a family of stability regions of multistep methods. Here, we show that the same algebraic approach allows us to replace the sector with more general shapes: we use again the multistep family (18) as a test example and determine the optimal stability region that contains the largest parabola. The motivation for considering the shape of a parabola comes from [18] (“for advection-diffusion equations, stability within a parabola<sup>2</sup> can be more relevant than for a wedge”), or from [5, Sections 3–4] (where linearly implicit Runge–Kutta methods are developed for the numerical integration of semidiscrete equations originating from spatial discretizations of PDEs of advection-reaction-diffusion type).

With  $P_{\beta_1, \beta_0}$  and  $\mathcal{S}_{\beta_1, \beta_0}$  defined in (20)–(21), we are now looking for the largest possible  $m > 0$  such that the stability region of a suitable member of the family (18) contains the parabola

$$\mathcal{P}_m := \{(\xi, \eta) \in \mathbb{R}^2 : \xi \leq 0, \eta \in \mathbb{R}, \eta^2 \leq m|\xi|\}, \quad (24)$$

that is, the inclusion

$$\mathcal{P}_m \subset \mathcal{S}_{\beta_1, \beta_0} \quad (25)$$

holds. Clearly, we need  $\overset{\circ}{A}$ -stability again to have (25) with some  $m > 0$ , so from now on, by Lemma 5.1, we can assume that  $(\beta_1, \beta_0) \in W$  (see Fig. 9).

<sup>2</sup>Similarly to Footnote 1, an analogous typo is present in [18, Section 2.3] when the notion of “stability within a parabola” is defined. There we should have again  $\tan$  instead of  $\arctan$ , that is,

$$\xi \leq 0, \quad |\eta^2/\xi| \leq \tan(\beta) \quad \text{with some angle } 0 < \beta \leq \pi/2.$$



In Appendix B.1, we apply a simple geometric argument: we first formulate the RLCs for the members of the multistep family as implicit curves  $\{(\xi, \eta) \in \mathbb{R}^2 : F_{\beta_1, \beta_0}(\xi, \eta) = 0\}$ , then invoke the notion of discriminant [14] to construct a polynomial in  $m$  (and depending on the parameters  $\beta_1$  and  $\beta_0$ ) whose suitable root can yield the optimal value  $\tilde{m}_{\text{opt}}$  in (25). The simple observation is the same as the one used in Section 3.1 (or in Section 4): the optimal inscribed object (now a parabola) touches the boundary of the optimal stability region.

Based on this technique and by using *Mathematica*, we conjecture that the parameter values  $\beta_1 = 1/5$  and  $\beta_0 = 37/40$  give  $\tilde{m}_{\text{opt}} = 6/5$ . In Appendix B.2, we use a uniqueness argument to rigorously prove this conjecture. We emphasize that, similarly to Appendix A.2, no RLCs are involved in this uniqueness proof; the RLCs are used only as auxiliary objects to conjecture the optimum. Given the complexity of intermediate calculations, it is again surprising that the final result  $\tilde{m}_{\text{opt}}$  is a simple rational number. In summary, we have the following theorem.

**Theorem 6.1** *Suppose that  $(\beta_1, \beta_0) \in W$ . Then the largest  $m > 0$  such that (25) holds is  $m \equiv \tilde{m}_{\text{opt}} := 6/5$ .*

**Remark 6.2** The authors of [18] observe that “for the methods considered in this paper, a large angle  $\alpha$  will correspond to a large  $\beta$ ” (with  $\alpha$  and  $\beta$  interpreted in our Footnotes 1 and 2). According to our results, the optimal  $(\beta_1, \beta_0)$  parameter pairs  $(3/8, 3/4)$  and  $(1/5, 37/40)$ —determining the stability regions with the largest inscribed sector and parabola, respectively—do not coincide, although they are both located on the right boundary of  $W$  in Fig. 9 (see also Remark B.1).

**Acknowledgment** Open access funding provided by Eötvös Loránd University (ELTE). The author would like to thank the anonymous referees for their suggestions, which have improved the presentation of the material.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix A

### A.1 Proof of Lemma 5.1

*Proof* Let us fix some  $(\beta_1, \beta_0) \in \mathbb{R}^2$ . For  $\xi < 0$ ,  $P_{\beta_1, \beta_0}(\zeta, \xi, 0) = 0$  is equivalent to  $\text{LHS}(\zeta) = \text{RHS}(\zeta)$  with

$$\text{LHS}_{\beta_1, \beta_0}(\zeta) := \beta_0 \zeta^3 + \beta_1 \zeta^2 - (3\beta_0 + 2\beta_1 - 3)\zeta + 2\beta_0 + \beta_1 - \frac{3}{2}$$

and  $\text{RHS}_{\xi}(\zeta) := (\zeta^3 - 3\zeta^2/4 - 1/4)/\xi$ . Clearly, if  $|\xi|$  is large enough, the coefficients of the RHS polynomial can be arbitrarily close to 0. So by the fact that the roots of a polynomial are continuous functions of its coefficients, we get that “the  $\zeta_j$

roots of  $\text{LHS}(\zeta) = \text{RHS}(\zeta)$  can be made arbitrarily close to those of  $\text{LHS}(\zeta) = 0$  by choosing  $|\xi|$  large." To make the previous "statement" precise, we distinguish two cases according to whether the leading coefficient of LHS vanishes or not: for  $\beta_0 = 0$ , the LHS polynomial has at most two roots, whereas the difference  $\text{LHS} - \text{RHS}$  has three.

Case I:  $\beta_0 \neq 0$ . By the above statement we easily see that if  $\text{LHS}_{\beta_1, \beta_0}(\cdot) \notin \mathbf{vN}$ , then  $P_{\beta_1, \beta_0}(\cdot, \xi, 0) \notin \mathbf{svN}$  for  $|\xi|$  large enough. We now show that

$$(\beta_1, \beta_0) \notin W \implies \text{LHS}_{\beta_1, \beta_0}(\cdot) \notin \mathbf{vN}. \quad (26)$$

So let us suppose in the rest of Case I that  $(\beta_1, \beta_0) \notin W$  and  $\beta_0 \neq 0$ .

Case I<sub>a</sub>. First, we check the case when  $\text{cc LHS}_{\beta_1, \beta_0}(\cdot) = 0$ . Then

$$\text{LHS}_{\beta_1, \beta_0}(\zeta) = \zeta/4 \left[ (2\beta_1 - 3)\zeta^2 - 4\beta_1\zeta + 2\beta_1 - 3 \right],$$

and, since now  $2\beta_1 - 3 \neq 0$ , we can apply Theorem 2.4 to the above polynomial in  $[\cdot \cdot \cdot]$ : due to  $[\cdot \cdot \cdot]^r \equiv 0$  we have that  $[\cdot \cdot \cdot] \in \mathbf{vN}$  if and only if  $\zeta \mapsto [\cdot \cdot \cdot]' = 2(2\beta_1 - 3)\zeta - 4\beta_1 \in \mathbf{vN}$ . But we directly see that this last linear polynomial  $\notin \mathbf{vN}$ , because  $(\beta_1, \beta_0) \notin W$  and  $\text{cc LHS}_{\beta_1, \beta_0}(\cdot) = 0$  imply  $\beta_1 > 3/4$ .

Case I<sub>b</sub>. The conditions  $\text{cc LHS}_{\beta_1, \beta_0}(\cdot) \neq 0 \neq \beta_0$  mean that we can apply Theorem 2.4 to  $\text{LHS}_{\beta_1, \beta_0}(\cdot)$ . It is easy to verify that  $(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r$  does not vanish identically, so  $\text{LHS}_{\beta_1, \beta_0}(\cdot) \in \mathbf{vN}$  if and only if

$$|\text{lc LHS}_{\beta_1, \beta_0}(\cdot)| > |\text{cc LHS}_{\beta_1, \beta_0}(\cdot)| \text{ and } (\text{LHS}_{\beta_1, \beta_0}(\cdot))^r \in \mathbf{vN}. \quad (27)$$

We show in Cases I<sub>b1</sub> and I<sub>b2</sub> below that (27) never occurs. First, we observe that the inequality constraint in (27) yields that  $\text{lc}(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r \neq 0$ .

Case I<sub>b1</sub>. If  $\text{cc}(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r = 0$ , then the polynomial  $(\text{LHS}_{\beta_1, \beta_0}(\zeta))^r$  has exactly two roots:  $\zeta_1 = 0$  and

$$\zeta_2 = 2 - \frac{3}{2(6\beta_0 + 2\beta_1 - 3)} + \frac{3}{2(2\beta_0 + 2\beta_1 - 3)}.$$

One directly checks that  $\text{cc}(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r = 0$  and  $(\beta_1, \beta_0) \notin W$  imply  $|\zeta_2| > 1$ .

Case I<sub>b2</sub>. If  $\text{cc}(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r \neq 0$ , we apply Theorem 2.4 to get that the quadratic polynomial  $(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r \in \mathbf{vN}$  if and only if either Case I<sub>b2 $\alpha$</sub>  or I<sub>b2 $\beta$</sub>  below occurs.

Case I<sub>b2 $\alpha$</sub> : when  $(\text{LHS}_{\beta_1, \beta_0}(\cdot))^{rr} \equiv 0$  and  $[(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r]' \in \mathbf{vN}$ . In this case, however, the unique root of the polynomial  $[(\text{LHS}_{\beta_1, \beta_0}(\cdot))^r]'$ ,

$$\zeta_1 = 1 - \frac{3}{4(6\beta_0 + 2\beta_1 - 3)} + \frac{3}{4(2\beta_0 + 2\beta_1 - 3)},$$

has absolute value  $> 1$ .

Case  $I_{b2\beta}$ : when  $|\text{lc}(\text{LHS}_{\beta_1, \beta_0}(\cdot))^{\mathbf{r}}| > |\text{cc}(\text{LHS}_{\beta_1, \beta_0}(\cdot))^{\mathbf{r}}|$  and  $(\text{LHS}_{\beta_1, \beta_0}(\cdot))^{\mathbf{r}\mathbf{r}} \in \mathbf{vN}$ . But then the unique root of  $(\text{LHS}_{\beta_1, \beta_0}(\cdot))^{\mathbf{r}\mathbf{r}}$  is

$$\zeta_1 = 1 - \frac{3(2\beta_0 + 2\beta_1 - 3)}{24\beta_0^2 + 32\beta_1\beta_0 - 36\beta_0 + 8\beta_1^2 - 18\beta_1 + 9},$$

for which we again have  $|\zeta_1| > 1$ , completing Case I.

Case II:  $\beta_0 = 0$ . Then

$$P_{\beta_1, 0}(\zeta, \xi, 0) = \zeta^3 - \zeta^2 \left( \beta_1 \xi + \frac{3}{4} \right) - (3 - 2\beta_1) \xi \zeta - \left( \beta_1 \xi - \frac{3\xi}{2} + \frac{1}{4} \right),$$

and the leading coefficient of this cubic polynomial is 1. For each fixed  $\beta_1 \in \mathbb{R}$  we see that at least one of its coefficients is unbounded as  $\xi \rightarrow -\infty$ , so (by Vieta's formulae) at least one of its roots  $\zeta(\xi)$  is unbounded as  $\xi \rightarrow -\infty$ . Hence,  $(-\infty, 0) \times \{0\} \subset \mathcal{S}_{\beta_1, 0}$  cannot hold.  $\square$

## A.2 Proof of Theorem 5.3

*Proof* In the proof, we suppose  $m > 0$  and, due to Lemma 5.1, that  $(\beta_1, \beta_0) \in W$ .

**Step 1.** Let us apply the same ideas as in Section A.1 but along the ray  $\eta = -m\xi$ . For  $\xi < 0$ , we consider the roots of  $P_{\beta_1, \beta_0}(\cdot, \xi, -m\xi)$  and get that

$$\text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \notin \mathbf{vN} \implies P_{\beta_1, \beta_0}(\cdot, \xi, -m\xi) \notin \mathbf{svN}$$

for some  $|\xi|$  large enough, where the corresponding “modified left-hand side” is defined as

$$\text{MLHS}_{\beta_1, \beta_0, m}(\zeta) := \text{LHS}_{\beta_1, \beta_0}(\zeta) - \frac{3}{2}im\zeta^2,$$

and we have also taken into account that  $\text{lc MLHS}_{\beta_1, \beta_0, m}(\cdot) = \beta_0 \neq 0$ . (The corresponding “modified right-hand side” would be the same  $\text{RHS}_{\xi}(\zeta)$  as in Section A.1.) Hence, if the inclusion (22) holds with some  $m > 0$ , then  $\text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \in \mathbf{vN}$ .

**Step 2.** In this step, we derive a necessary condition for  $\text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \in \mathbf{vN}$ . First, one simply checks via Theorem 2.4 that  $\text{cc MLHS}_{\beta_1, \beta_0, m}(\cdot) = 0$ ,  $\text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \in \mathbf{vN}$  and  $m > 0$  cannot be simultaneously true. So we can suppose

$$\text{lc MLHS}_{\beta_1, \beta_0, m}(\cdot) \neq 0 \neq \text{cc MLHS}_{\beta_1, \beta_0, m}(\cdot).$$

We check that  $(\text{MLHS}_{\beta_1, \beta_0, m}(\cdot))^{\mathbf{r}}$  does not vanish identically, and that

$$|\text{lc MLHS}_{\beta_1, \beta_0, m}(\cdot)| > |\text{cc MLHS}_{\beta_1, \beta_0, m}(\cdot)|.$$

Then by Theorem 2.4 we have that

$$\text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \in \mathbf{vN} \iff (\text{MLHS}_{\beta_1, \beta_0, m}(\cdot))^{\mathbf{r}} \in \mathbf{vN}.$$

Now we see that

$$\text{lc} \left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{r}} \neq 0 \neq \text{cc} \left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{r}},$$

and  $\left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{rr}}$  does not vanish identically. Thus, Theorem 2.4 yields that

$$\left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{r}} \in \mathbf{vN}$$

if and only if

$$\left| \text{lc} \left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{r}} \right| > \left| \text{cc} \left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{r}} \right| \quad (28)$$

and

$$\left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{rr}} \in \mathbf{vN}. \quad (29)$$

Clearly,  $\deg \left( \text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \right)^{\mathbf{rr}} \leq 1$ , and we directly confirm that (28) implies that the degree is exactly 1. From this, we obtain that (28) and (29) hold if and only if (28) and

$$\left| 1 + im + \frac{2im\beta_0}{4\beta_0 + 2\beta_1 - 3} - \frac{3(1 + im) \left[ \beta_0 (6m^2 + 2) + (2\beta_1 - 3)(m^2 + 1) \right]}{24\beta_0^2 + 4\beta_0(8\beta_1 + 3m^2 - 9) + (2\beta_1 - 3)(4\beta_1 + 3m^2 - 3)} \right| \leq 1 \quad (30)$$

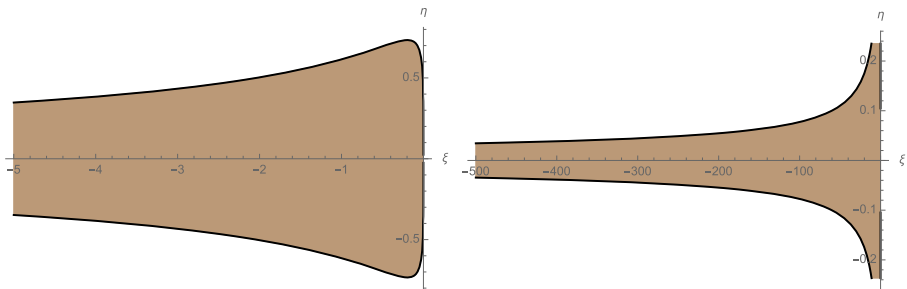
hold. In particular, (28) guarantees that the denominators appearing in (30) are non-zero; hence, from now on, we can restrict the parameters  $(\beta_1, \beta_0) \in W$  to the set  $(\beta_1, \beta_0) \in W \setminus L$  with

$$L := \left\{ (\beta_1, \beta_0) \in \mathbb{R}^2 : \beta_0 = \frac{3 - 2\beta_1}{4} \right\}, \quad (31)$$

being the left edge of the wedge  $W$  (see Fig. 10).

By defining

$$\begin{aligned} C_4 &:= -9(4\beta_0 + 2\beta_1 - 3)^2, \\ C_2 &:= 2 \left[ 864\beta_0^4 + 864(2\beta_1 - 3)\beta_0^3 + 288(4\beta_1^2 - 13\beta_1 + 10)\beta_0^2 + \right. \\ &\quad \left. 4(80\beta_1^3 - 420\beta_1^2 + 684\beta_1 - 351)\beta_0 + (3 - 2\beta_1)^2(8\beta_1^2 - 36\beta_1 + 27) \right], \end{aligned}$$



**Fig. 10** These figures show the stability region  $S_{3/4, 3/8}$  corresponding to the method with  $(\beta_1, \beta_0) = (3/4, 3/8)$  (i.e., the vertex of the wedge in Fig. 9). Such methods with  $(\beta_1, \beta_0) \in W \cap L$  (see (31)) are  $\mathcal{A}$ -stable, but  $\mathcal{A}_m \subset \mathcal{S}_{\beta_1, \beta_0}$  (see (22)) does not hold with any  $m > 0$

$$C_0 := -3(4\beta_0 + 2\beta_1 - 3)^2(8\beta_0 + 8\beta_1 - 9)$$

and

$$Q_{\beta_1, \beta_0}(m) := C_4 m^4 + C_2 m^2 + C_0,$$

it is easily verified after some factorization and simplification that

$$(28) \text{ and } (30) \iff (28) \text{ and } Q_{\beta_1, \beta_0}(m) \geq 0.$$

In particular,  $\text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \in \mathbf{vN}$  implies  $Q_{\beta_1, \beta_0}(m) \geq 0$ .

**Step 3.** We see that  $C_4 < 0$  and  $C_0 \geq 0$  for  $(\beta_1, \beta_0) \in W \setminus L$ ; hence, we can denote the largest real root of the polynomial  $Q_{\beta_1, \beta_0}(\cdot)$  by  $m^*(\beta_1, \beta_0) \in [0, +\infty)$ . Consequently, if  $\text{MLHS}_{\beta_1, \beta_0, m}(\cdot) \in \mathbf{vN}$ , then  $m \leq m^*(\beta_1, \beta_0)$ . We now conjecture (by using *Mathematica*'s `Maximize` command, for example) that

$$m^*(\beta_1, \beta_0) \leq \frac{1}{2} \quad \text{for } (\beta_1, \beta_0) \in W \setminus L, \quad (32)$$

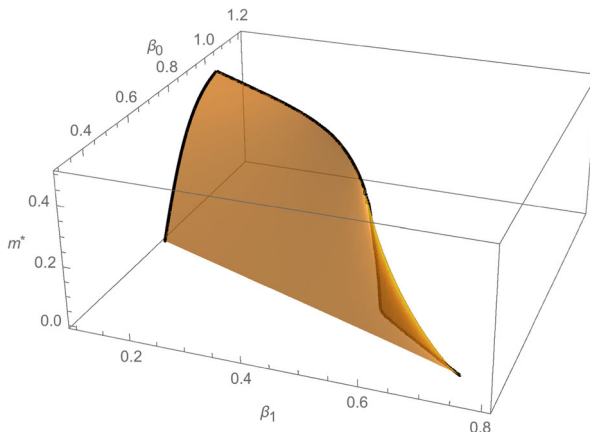
and  $m^*(\beta_1, \beta_0) = 1/2$  occurs precisely for  $(\beta_1, \beta_0) = (3/8, 3/4)$  (see Fig. 11). With this conjectured optimal  $m^*$  value, we can prove (32) and the uniqueness property in an elementary way.

By introducing the shifted variable  $M := m - 1/2$ , we rewrite  $Q_{\beta_1, \beta_0}(m)$  as

$$\sum_{j=0}^4 \widehat{C}_j(\beta_1, \beta_0) M^j. \quad (33)$$

Then we check that

$$(\beta_1, \beta_0) \in W \setminus L \implies \widehat{C}_j(\beta_1, \beta_0) < 0 \text{ for } 1 \leq j \leq 4.$$



**Fig. 11** The function  $m^*$  defined in Step 3 in Appendix A.2. Its maximum value is located at  $(\beta_1, \beta_0, m^*) = (3/8, 3/4, 1/2)$

Moreover, we have

$$\begin{aligned}\widehat{C}_0(\beta_1, \beta_0) \equiv & 6912\beta_0^4 + 768(18\beta_1 - 35)\beta_0^3 \\ & + 48(192\beta_1^2 - 880\beta_1 + 813)\beta_0^2 \\ & + 40(64\beta_1^3 - 528\beta_1^2 + 1062\beta_1 - 621)\beta_0 \\ & + (3 - 2\beta_1)^2(64\beta_1^2 - 672\beta_1 + 639),\end{aligned}$$

$$(\beta_1, \beta_0) \in W \setminus L \implies \widehat{C}_0(\beta_1, \beta_0) \leq 0$$

and

$$[(\beta_1, \beta_0) \in W \setminus L \text{ and } \widehat{C}_0(\beta_1, \beta_0) = 0] \iff (\beta_1, \beta_0) = (3/8, 3/4).$$

On the one hand, these mean that (33) is negative for  $M > 0$  and  $(\beta_1, \beta_0) \in W \setminus L$ . On the other hand, for  $M = 0$  the polynomial (33) is zero if and only if  $(\beta_1, \beta_0) = (3/8, 3/4)$ .

Therefore, we have proved that if (22) holds with some  $m > 0$ , then  $m \leq 1/2$ , and if  $m = 1/2$  is possible at all, then  $(\beta_1, \beta_0) = (3/8, 3/4)$ .

**Step 4.** In this final step we show that  $m = 1/2$  in (22) can be achieved, by showing that  $\mathcal{A}_{1/2} \subset \mathcal{S}_{3/8, 3/4}$ , that is,

$$(\xi, \eta) \in \mathcal{A}_{1/2} \implies P_{3/8, 3/4}(\cdot, \xi, \eta) \in \mathbf{svN}. \quad (34)$$

Let us fix such a pair  $(\xi, \eta)$ . One sees that

$$|\mathrm{lc} P_{3/8, 3/4}(\cdot, \xi, \eta)| > |\mathrm{cc} P_{3/8, 3/4}(\cdot, \xi, \eta)|,$$

and in the  $\mathrm{cc} P_{3/8, 3/4}(\cdot, \xi, \eta) = 0$  case (34) is easily verified to hold. Otherwise, if  $\mathrm{cc} \neq 0$ , we check that  $(P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{r}}$  does not vanish identically; so by Theorem 2.5, we have that  $P_{3/8, 3/4}(\cdot, \xi, \eta) \in \mathbf{svN}$  if and only if

$$(P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{r}} \in \mathbf{svN}. \quad (35)$$

We have that  $\mathrm{lc} (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{r}} \neq 0$ . Moreover,  $\mathrm{cc} (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{r}} = 0$  for  $\xi = -2$  or  $\xi = -2/3$ , in which cases (35) holds. So we can suppose from now on that  $\mathrm{cc} (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{r}} \neq 0$ . Then one proves that

$$\begin{aligned}\mathrm{lc} (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{rr}} = \\ -\frac{81\eta^2\xi^2}{256} - \frac{27\eta^2\xi}{64} - \frac{9\eta^2}{64} + \frac{81\xi^4}{512} - \frac{783\xi^3}{512} + \frac{441\xi^2}{128} - \frac{423\xi}{128} + \frac{27}{32} \neq 0,\end{aligned}$$

so  $\deg (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{rr}} = 1$ . Hence, by using Theorem 2.5 again, we get that (35) holds if and only if

$$|\mathrm{lc} (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{r}}| > |\mathrm{cc} (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{r}}| \text{ and } (P_{3/8, 3/4}(\cdot, \xi, \eta))^{\mathbf{rr}} \in \mathbf{svN}$$

hold. Finally, we check that these last two conditions are satisfied for any  $(\xi, \eta) \in \mathcal{A}_{1/2}$  pair not excluded earlier during the case separations.  $\square$

**Remark A.1** By defining

$$F_{\text{opt}}(\xi, \eta) := 12\eta^4(3\xi + 2)^2 -$$

$$3\eta^2\xi \left( 9\xi^3 + 192\xi^2 - 620\xi + 368 \right) + 16\xi \left( 3\xi^2 - 7\xi + 6 \right)^2$$

and applying Theorem 2.5, it is straightforward to show (cf. Step 4 in the above proof) that

$$\mathcal{S}_{3/8,3/4} = \{(\xi, \eta) \in \mathbb{R}^2 : \xi \leq 0, \eta \in \mathbb{R}, F_{\text{opt}}(\xi, \eta) \leq 0\},$$

see Fig. 12 and cf. Remark B.1.

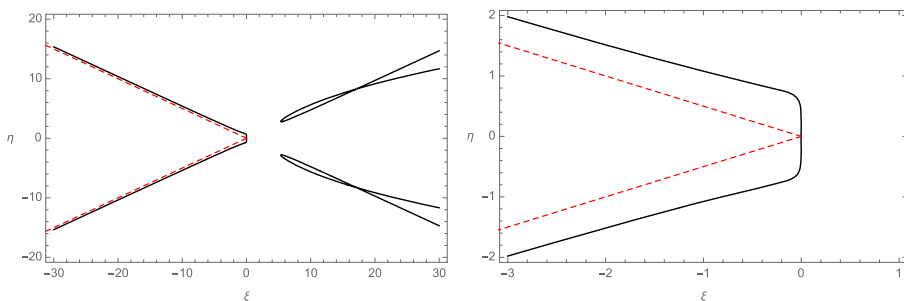
## Appendix B

### B.1 Locating the candidate optimum for Theorem 6.1

For a given  $(\beta_1, \beta_0) \in W$  pair, we can represent the RLC of the corresponding multistep method of the family (18) as an implicit curve of the form

$$\{(\xi, \eta) \in \mathbb{R}^2 : \xi \leq 0, \eta \in \mathbb{R}, F_{\beta_1, \beta_0}(\xi, \eta) = 0\} \quad (36)$$

by using the transformations in Section 3.1 as follows. First, we perform the substitution  $\zeta \mapsto \frac{i-t}{i+t}$  in the polynomial (20), then eliminate  $t \in \mathbb{R}$  by taking the resultant of the real and imaginary parts of  $P_{\beta_1, \beta_0} \left( \frac{i-t}{i+t}, \xi, \eta \right)$ . The resulting polynomial can be factored to get  $2^{34} \cdot 9 \cdot (1 - \beta_0\xi)^6 \cdot F_{\beta_1, \beta_0}(\xi, \eta)$ ; the normalization with  $F_{\beta_1, \beta_0}(0, 1) = 9$  has been used to make this polynomial  $F_{\beta_1, \beta_0}$  unique. The term  $(1 - \beta_0\xi)^6$  (cf. the leading coefficient of  $P_{\beta_1, \beta_0}(\cdot, \xi, \eta)$ ) does not vanish now due to  $\xi \leq 0$  and  $(\beta_1, \beta_0) \in W$ ; hence, (36) is obtained. We are not going to display the



**Fig. 12** The implicit curve  $\{(\xi, \eta) \in \mathbb{R}^2 : F_{\text{opt}}(\xi, \eta) = 0\}$  (see Remark A.1), being the boundary of the optimal stability region  $\mathcal{S}_{3/8,3/4}$  in the left half-plane, is shown in the left figure and a close-up in the right figure in black. The dashed red lines represent the boundary of the largest infinite sector  $\mathcal{A}_{1/2}$  that can be included in the stability region

polynomial  $F_{\beta_1, \beta_0}(\xi, \eta)$  explicitly: it contains 82 terms in its expanded form and its degree in the variables/parameters  $(\xi, \eta, \beta_1, \beta_0)$  is  $(6, 4, 4, 4)$ .

Now supposing that the RLC (36) describes the boundary of the stability region of the multistep method determined by the given pair  $(\beta_1, \beta_0)$ , it is reasonable to expect that, say, the upper branch of the largest parabola inscribed in  $\mathcal{S}_{\beta_1, \beta_0}$ ,  $\{(\xi, \eta) \in \mathbb{R}^2 : \xi < 0, \eta > 0, \eta^2 = -m\xi\}$ , touches the RLC (36) at some finite point. In this case, the polynomial

$$(-\infty, 0) \ni \xi \mapsto F_{\beta_1, \beta_0}(\xi, \sqrt{-m\xi})$$

has a multiple root there—it is indeed a polynomial, because in our situation  $F_{\beta_1, \beta_0}(\xi, \eta)$  contains only even powers of  $\eta$  (namely,  $\eta^2$  and  $\eta^4$ ). Moreover, we now have

$$F_{\beta_1, \beta_0}(\xi, \sqrt{-m\xi}) = \xi \cdot \tilde{Q}_{\beta_1, \beta_0, m}(\xi),$$

where  $\tilde{Q}_{\beta_1, \beta_0, m}(\cdot)$  is a quartic polynomial. The existence of a multiple root of  $\tilde{Q}_{\beta_1, \beta_0, m}(\cdot)$  implies that the discriminant of this polynomial (with respect to  $\xi$ ), denoted by  $\tilde{\Delta}_{\beta_1, \beta_0}(m)$ , vanishes. *Mathematica* yields that

$$\tilde{\Delta}_{\beta_1, \beta_0}(m) = -2^{13} \cdot 3^6 \cdot m^2 (9\beta_0 + 4\beta_1 - 6)^2 \times$$

$$(64\beta_1^4 m^3 + \dots - 4410)^2 (590976 \beta_0^2 \beta_1^3 m^5 + \dots - 24402696417),$$

where the “ $\dots$ ” symbols contain 57 and 228 terms, respectively. We see that the factor  $9\beta_0 + 4\beta_1 - 6$  above is always positive in  $W$ . In this way, we can determine the parameter  $m$  of the largest parabola within the region bounded by the RLC for any fixed  $(\beta_1, \beta_0) \in W$ .

**Remark B.1** By setting  $(\beta_1, \beta_0) = (3/8, 3/4)$  for example (corresponding to the “sector-optimal” method in Section A.2), we have that the RLC in (36) is identical to  $3/16 \cdot F_{\text{opt}}(\xi, \eta)$  in Remark A.1, implying that the RLC in the left half-plane  $\xi \leq 0$  represents the boundary of the stability region  $\mathcal{S}_{3/8, 3/4}$ . Now,  $\tilde{\Delta}_{3/8, 3/4}(m)$  can be written as

$$-\frac{3^{23}}{2^{13}} (3m - 1)^2 m^2 (m + 4)^4 (3m + 16)^2 (36m^3 + 1362m^2 + 343m - 2116),$$

from which we can prove that the largest parabola  $\mathcal{P}_m$  contained in  $\mathcal{S}_{3/8, 3/4}$  has  $m \approx 1.11226$  (being the unique positive root of the polynomial  $\{36, 1362, 343, -2116\}$ ).

By studying the positive roots of  $\tilde{\Delta}_{\beta_1, \beta_0}(\cdot)$  as  $(\beta_1, \beta_0)$  varies within  $W$ , we can conjecture that the value of  $m$  in (25) cannot be greater than  $6/5$  for the family (18). Moreover,  $m = 6/5$  occurs only for  $\beta_1 = 1/5$  and  $\beta_0 = 37/40$ , and in this case the RLC and the upper parabola branch touch each other at  $(\xi, \eta) = (-10/7, 2\sqrt{3/7})$ . Since the polynomial  $\tilde{\Delta}_{\beta_1, \beta_0}(m)$  is much more complicated than the corresponding polynomial  $Q_{\beta_1, \beta_0}(m)$  in Appendix A.2, this time *Mathematica* could not confirm in a reasonable amount of computing time that the value  $m = 6/5$  is indeed the optimal one.



## B.2 Proof of optimality in Theorem 6.1

However, once the unique optimum has been conjectured properly, the proof of optimality becomes straightforward to complete.

**Step 1.** By assuming  $(\beta_1, \beta_0) \in W$  throughout the step, we show that the point  $(\xi_0, \eta_0) := (-10/7, 2\sqrt{3/7})$  belongs to precisely one stability region in the family, by verifying that

$$P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0) \in \mathbf{svN} \iff (\beta_1, \beta_0) = (1/5, 37/40).$$

To see this, first we check that  $\text{lc } P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0) \neq 0$ . Moreover, it is easily seen that  $\text{cc } P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0)$  vanishes exactly for  $\beta_1 \leq 23/40$  and  $\beta_0 = (67 - 40\beta_1)/80$ , and in this case the polynomial  $P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0)/\xi$  has  $\deg = 2$  but  $\notin \mathbf{svN}$ , as a recursive application of Theorem 2.5 shows. Then we can also prove that  $(P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{r}}$  does not vanish identically, and that

$$|\text{lc } P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0)| > |\text{cc } P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0)|.$$

Thus, according to Theorem 2.5,

$$P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0) \in \mathbf{svN} \iff (P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{r}} \in \mathbf{svN}.$$

Now we repeat the above process with  $(P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{r}}$ . We prove that

$$|\text{lc } (P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{r}}| > |\text{cc } (P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{r}}| > 0$$

and that  $(P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{rr}}$  does not vanish identically, so by Theorem 2.5 we have that

$$(P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{r}} \in \mathbf{svN} \iff (P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{rr}} \in \mathbf{svN}.$$

But  $(P_{\beta_1, \beta_0}(\cdot, \xi_0, \eta_0))^{\mathbf{rr}}$  is a linear polynomial (it is easily checked that it cannot be a constant polynomial), so its unique (complex) root can be directly expressed: one sees that the absolute value of this root is  $\leq 1$  if and only if

$$\begin{aligned} & (8\beta_0 + 8\beta_1 - 19)(120\beta_0 + 40\beta_1 - 39) \left( 483840000\beta_0^4 + 967680000\beta_1\beta_0^3 - \right. \\ & 1989440000\beta_0^3 + 645120000\beta_1^2\beta_0^2 - 2967744000\beta_1\beta_0^2 + 2890070400\beta_0^2 + \\ & 179200000\beta_1^3\beta_0 - 1404096000\beta_1^2\beta_0 + 2856374400\beta_1\beta_0 - 1693045320\beta_0 + \\ & \left. 17920000\beta_1^4 - 214336000\beta_1^3 + 673766400\beta_1^2 - 792582600\beta_1 + 301631887 \right) \leq 0. \end{aligned}$$

The product of the first two factors is strictly negative in  $W$ , and a standard constrained optimization computation shows that the third factor is  $\geq 0$  in  $W$  if and only if  $(\beta_1, \beta_0) = (1/5, 37/40)$ , completing Step 1.

**Step 2.** Since  $\mathcal{P}_{m_1} \subseteq \mathcal{P}_{m_2}$  is equivalent to  $0 < m_1 \leq m_2$  (see (24)), and now  $|\eta_0^2/\xi_0| = 6/5$ , the uniqueness property in the previous step implies that  $m \geq 6/5$  in (25) can hold only for  $(\beta_1, \beta_0) = (1/5, 37/40)$ . In this step, we

verify that (25) indeed holds with  $m = 6/5$  and  $(\beta_1, \beta_0) = (1/5, 37/40)$ ; that is, we show that  $P_{1/5, 37/40}(\cdot, \xi, \eta) \in \mathbf{svN}$  for any  $(\xi, \eta) \in \mathcal{P}_{6/5}$ .

Let us pick and fix an arbitrary point  $(\xi, \eta) \in \mathcal{P}_{6/5}$ . Then we easily see that

$$|\mathrm{lc} P_{1/5, 37/40}(\cdot, \xi, \eta)| > |\mathrm{cc} P_{1/5, 37/40}(\cdot, \xi, \eta)|,$$

and this  $\mathrm{cc} = 0$  if and only if  $\xi = -5/11$ ; in this case, Theorem 2.5 tells us that  $\zeta \mapsto P_{1/5, 37/40}(\zeta, -5/11, \eta) = \zeta(125\zeta^2 - 132i\zeta\eta - 58\zeta - 7)/88 \in \mathbf{svN}$ . So for  $\xi \neq -5/11$ , again by Theorem 2.5 we get that

$$P_{1/5, 37/40}(\cdot, \xi, \eta) \in \mathbf{svN} \iff (P_{1/5, 37/40}(\cdot, \xi, \eta))^{\mathbf{r}} \in \mathbf{svN},$$

provided that  $(P_{1/5, 37/40}(\cdot, \xi, \eta))^{\mathbf{r}}$  does not vanish identically. But this non-vanishing condition is true because

$$|\mathrm{lc} (P_{1/5, 37/40}(\cdot, \xi, \eta))^{\mathbf{r}}| > |\mathrm{cc} (P_{1/5, 37/40}(\cdot, \xi, \eta))^{\mathbf{r}}| > 0.$$

Moreover, since  $|\mathrm{lc} (P_{1/5, 37/40}(\cdot, \xi, \eta))^{\mathbf{rr}}|$  is also positive, the above with Theorem 2.5 imply that

$$P_{1/5, 37/40}(\cdot, \xi, \eta) \in \mathbf{svN} \iff (P_{1/5, 37/40}(\cdot, \xi, \eta))^{\mathbf{rr}} \in \mathbf{svN}.$$

The positivity of  $|\mathrm{lc} (\cdot \cdot \cdot)^{\mathbf{rr}}|$  yields that  $(P_{1/5, 37/40}(\cdot, \xi, \eta))^{\mathbf{rr}}$ , a  $\deg = 1$  polynomial, has a unique root. The absolute value of this (real or complex) root is  $\leq 1$  if and only if  $(3\xi - 10)(59\xi - 30) \cdot \tilde{F}_{\mathrm{opt}}(\xi, \eta) \leq 0$ , where

$$\tilde{F}_{\mathrm{opt}}(\xi, \eta) := 720\eta^4(11\xi + 5)^2 -$$

$$\eta^2\xi \left( 19575\xi^3 + 485696\xi^2 - 1009140\xi + 464400 \right) + 240\xi \left( 22\xi^2 - 49\xi + 30 \right)^2.$$

Now,  $(3\xi - 10)(59\xi - 30) > 0$ , and one checks that  $\tilde{F}_{\mathrm{opt}}(\xi, \eta) \leq 0$  for  $(\xi, \eta) \in \mathcal{P}_{6/5}$ , completing Step 2.

**Step 3.** To complete the optimality proof, we finally show that

$$P_{1/5, 37/40}(\cdot, \xi_0, \eta_0 + \varepsilon) \notin \mathbf{svN} \text{ for any } \varepsilon \in (0, 1),$$

that is, we cannot have  $m > 6/5$  in (25). We repeat the same two-step reduction process as above and get that  $\varepsilon \in (0, 1)$  guarantees that

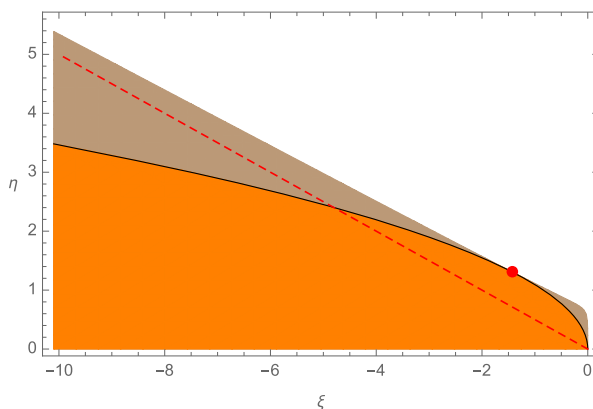
$$P_{1/5, 37/40}(\cdot, \xi_0, \eta_0 + \varepsilon) \in \mathbf{svN} \iff (P_{1/5, 37/40}(\cdot, \xi_0, \eta_0 + \varepsilon))^{\mathbf{rr}} \in \mathbf{svN}.$$

But this last  $(\cdot \cdot \cdot)^{\mathbf{rr}} \in \mathbf{svN}$  condition is equivalent to

$$\frac{1120\varepsilon \left( 27783\varepsilon^3 + 31752\sqrt{21}\varepsilon^2 + 1649620\varepsilon + 833776\sqrt{21} \right)}{\left( 1323\varepsilon^2 + 756\sqrt{21}\varepsilon - 50840 \right)^2} \leq 0,$$

so it cannot hold for any  $\varepsilon \in (0, 1)$ .

*Remark B.2* In addition to the inequality  $\tilde{F}_{\mathrm{opt}}(\xi, \eta) \leq 0$  in Step 2, we have that  $\tilde{F}_{\mathrm{opt}}(\xi, \eta) = 0$  for  $(\xi, \eta) \in \mathcal{P}_{6/5}$  if and only if  $(\xi, \eta) = (\xi_0, \pm\eta_0)$ . Moreover,



**Fig. 13** The figure shows the optimal stability region  $S_{1/5, 37/40}$  (brown) within the family (18) that contains the largest parabola  $\mathcal{P}_{6/5}$  (orange), see Theorem 6.1. The point  $(\xi_0, \eta_0) = (-10/7, 2\sqrt{3/7})$  is shown as a red dot. For comparison, the dashed red line from Fig. 12 is also included here

$\tilde{F}_{\text{opt}}(\xi, \eta) \equiv 2000 \cdot F_{1/5, 37/40}(\xi, \eta)$  (see (36)). On the other hand, by using the reduction process one can actually prove that

$$\{(\xi, \eta) \in \mathbb{R}^2 : \xi \leq 0, \eta \in \mathbb{R}, \tilde{F}_{\text{opt}}(\xi, \eta) \leq 0\} = S_{1/5, 37/40}.$$

These mean that the stability region boundary in the optimal case coincides with the corresponding RLC (in the left half-plane), and the boundary of the optimal inscribed parabola touches the stability region boundary in the open upper left half-plane at exactly one point, see Fig. 13 (and cf. Remarks 5.4 and B.1).

## References

1. Akrivis, G., Katsoprakis, E.: Maximum angles of  $A(\vartheta)$ -stability of backward difference formulae, submitted for publication
2. Bhattacharyya, S.P., Chapellat, H., Keel, L.H.: Robust Control: the Parametric Approach. Prentice Hall PTR (1995)
3. Bickart, T.A., Jury, E.I.: Arithmetic tests for  $A$ -stability,  $A[\alpha]$ -stability, and stiff-stability. BIT **18**(1), 9–21 (1978)
4. Butcher, J.C.: Numerical Methods for Ordinary Differential Equations. Wiley, Chichester (2008)
5. Calvo, M.P., de Frutos, J., Novo, J.: Linearly implicit Runge–Kutta methods for advection-reaction-diffusion equations. Appl. Numer. Math. **37**(4), 535–549 (2001)
6. Chakravarti, P.C., Kamel, M.S.: Stiffly stable second derivative multistep methods with higher order and improved stability regions. BIT **23**(1), 75–83 (1983)
7. Creedon, D.M., Miller, J.J.H.: The stability properties of  $q$ -step backward difference schemes. BIT **15**(3), 244–249 (1975)
8. Cryer, C.W.: On the instability of high order backward-difference multistep methods. BIT **12**, 17–25 (1972)
9. Cryer, C.W.: A new class of highly-stable methods:  $A_0$ -stable methods. BIT **13**, 153–159 (1973)
10. Enright, W.H.: Second derivative multistep methods for stiff ordinary differential equations. SIAM J. Numer. Anal. **11**(2), 321–331 (1974)
11. Friedli, A., Jeltsch, R.: An algebraic test for  $A_0$ -stability. BIT **18**(4), 402–414 (1978)

12. Gautschi, W.: Numerical Analysis. Springer Science+Business Media, New York (2012)
13. Gear, C.W.: Numerical Initial Value Problems in Ordinary Differential Equations. Prentice-Hall, Englewood Cliffs (1971)
14. Gelfand, I.M., Kapranov, M.M., Zelevinsky, A.V.: Discriminants, Resultants, and Multidimensional Determinants. Birkhäuser, Boston (1994)
15. Gottlieb, S., Ketcheson, D., Shu, C.-W.: Strong Stability Preserving Runge–Kutta and Multistep Time Discretizations. World Scientific Publishing Co., Hackensack (2011)
16. Hairer, E., Wanner, G.: On the instability of the BDF formulas. *SIAM J. Numer. Anal.* **20**(6), 1206–1209 (1983)
17. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems. Springer, Berlin (2002)
18. Hundsdorfer, W., Ruuth, S.J.: IMEX extensions of linear multistep methods with general monotonicity and boundedness properties. *J. Comput. Phys.* **225**(2), 2016–2042 (2007)
19. Hundsdorfer, W., Mozartova, A., Spijker, M.N.: Stepsize restrictions for boundedness and monotonicity of multistep methods. *J. Sci. Comput.* **50**(2), 265–286 (2012)
20. Jeltsch, R.: A necessary condition for  $A$ -stability of multistep multiderivative methods. *Math. Comp.* **30**(136), 739–746 (1976)
21. Jeltsch, R.: Note on  $A$ -stability of multistep multiderivative methods. *BIT* **16**(1), 74–78 (1976)
22. Jeltsch, R.: Stiff stability and its relation to  $A_0$ - and  $A(0)$ -stability. *SIAM J. Numer. Anal.* **13**(1), 8–17 (1976)
23. Jeltsch, R.: Stiff stability of multistep multiderivative methods. *SIAM J. Numer. Anal.* **14**(4), 760–772 (1977)
24. Jeltsch, R.: On the stability regions of multistep multiderivative methods. In: Numerical treatment of differential equations (Proc. Conf., Math. Forschungsinst., Oberwolfach, 1976), Lecture Notes in Math, vol. 631, pp. 63–80. Springer, Berlin (1978)
25. Jeltsch, R., Kratz, L.: On the stability properties of Brown’s multistep multiderivative methods. *Numer. Math.* **30**(1), 25–38 (1978)
26. Jeltsch, R.: Corrigendum: stiff stability of multistep multiderivative methods. *SIAM J. Numer. Anal.* **16**(2), 339–345 (1979)
27. Jeltsch, R., Nevanlinna, O.: Stability and accuracy of time discretizations for initial value problems. *Numer. Math.* **40**(2), 245–296 (1982)
28. Jeltsch, R.: Stability of time discretization, Hurwitz determinants and order stars. In: Stability theory (Ascona, 1995), *Internat. Ser. Numer. Math.*, vol. 121, pp. 191–204. Birkhäuser, Basel (1996)
29. Jeltsch, R., Torrilhon, M.: Essentially optimal explicit Runge–Kutta methods with application to hyperbolic-parabolic equations. *Numer. Math.* **106**(2), 303–334 (2007)
30. Ketcheson, D.I., Ahmadi, A.J.: Optimal stability polynomials for numerical integration of initial value problems. *Commun. Appl. Math. Comput. Sci.* **7**(2), 247–271 (2012)
31. Ketcheson, D.I., Kocsis, T.A., Lóczy, L.: On the absolute stability regions corresponding to partial sums of the exponential function. *IMA J. Numer. Anal.* **35**(3), 1426–1455 (2015)
32. Kirlinger, G.: Linear multistep methods applied to stiff initial value problems—a survey. *Math. Comput. Modelling* **40**(11–12), 1181–1192 (2004)
33. Lambert, J.D.: Computational Methods in Ordinary Differential Equations. Wiley, London (1973)
34. Liniger, W.: A criterion for  $A$ -stability of linear multistep integration formulae. *Computing* **3**, 280–285 (1968)
35. Lóczy, L.: Exact optimal values of step-size coefficients for boundedness of linear multistep methods. *Numer. Algor.* **77**(4), 1093–1116 (2018)
36. Marden, M.: Geometry of Polynomials, 2nd edn. Mathematical Surveys, No. 3. American Mathematical Society, Providence (1966)
37. Miller, J.J.H.: On the location of zeros of certain classes of polynomials with applications to numerical analysis. *J. Inst. Math. Appl.* **8**, 397–406 (1971)
38. Nørsett, S.P.: A criterion for  $A(\alpha)$ -stability of linear multistep methods. *BIT* **9**, 259–263 (1969)
39. de Oliveira, P., Patrício, F.: On fitting stability regions of second-derivative multi-step methods. *Comm. Appl. Numer. Methods* **8**(6), 351–360 (1992)
40. Sanz-Serna, J.M.: Some aspects of the boundary locus method. *BIT* **20**(1), 97–101 (1980)

41. Spijker, M.N.: The existence of stepsize-coefficients for boundedness of linear multistep methods. *Appl. Numer. Math.* **63**, 45–57 (2013)
42. Spijker, M.N.: Stability and boundedness in the numerical solution of initial value problems. *Math. Comp.* **86**(308), 2777–2798 (2017)
43. Spivak, M.: *Calculus*. Cambridge University Press, Cambridge (2006)
44. Süli, E., Mayers, D.F.: *An Introduction to Numerical Analysis*. Cambridge University Press, Cambridge (2003)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.